Modelling of Metabolic Pathways of *Saccharopolyspora erythraea* Using Flux Balance Analysis

A Thesis submitted to the University of London for the degree of

Doctor of Philosophy

By

Colin Mark Jaques

March 2004

The Advanced Centre for Biochemical Engineering
Department of Biochemical Engineering
University College London
Torrington Place
London WC1E 7JE
United Kingdom

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Soli Deo Gloria

Abstract

The objective of this thesis is to use metabolic modelling techniques to investigate primary and secondary metabolism in *S. erythraea* and from this to identify key factors controlling flux distribution during secondary metabolism.

S. erythraea is a member of the actinomycetes a group of bacteria responsible for the production of a number of commercially important small molecules. Actinomycete physiology is considerably more complicated than that seen in "simple" bacteria such as *E. coli*. The conjecture investigated in this thesis is that metabolic modelling techniques that take into account this extra complexity should be more useful in designing strategies for overproduction of desired metabolites than simpler models.

The thesis gives the first detailed description of the dynamic changes in biomass composition seen during the batch cultivation of *S. erythraea*. It further shows that incorporation of this information into a flux balance model of the organism's metabolism significantly improves the flux distributions generated especially in the stationary phase. Using this improved technique growth phase and stationary phase metabolism are investigated. Some of the unusual stationary phase behaviour is shown to be the result of glucose uptake being independent of demand. Rigid control of branch points in the metabolic network is not found suggesting that the organism's metabolism is flexible. A reverse metabolic engineering strategy is applied, two variants of the wild type organism are compared with an industrial strain. The industrial strain is found to have a considerably lower glucose uptake rate than the parental strain. The relationship between TCA cycle flux, oxidative phosphorylation and organic acid secretion is investigated using an uncoupler.

This project demonstrates that applied correctly flux balance analysis is a powerful tool for investigating actinomycete physiology. The insights gained are of direct relevance to the commercial production of secondary metabolites in *S. erythraea*.

Acknowledgements

It is impossible in this day and age to pursue cutting edge science and engineering without a large amount of help. I have received financial, material, intellectual and practical support from many sources and I wish to take this opportunity to thank those who supported me.

I am indebted to my supervisors, Yuhong Zhou, Frank Baganz and John Ward for setting up the opportunity for me to study *S. erythraea* in this way. All their time, thoughts and direction over the course of the project have proved invaluable. I also have much to be thankful for in the officemates I have had in the cloisters office over my time at UCL, with them I have batted back and forth many of the ideas developed in this thesis. Particular thanks go to Misti Ushio for help with modelling and physiology, to Preben Krabben for his willingness to share his huge knowledge of the biochemistry of actinomycetes and to help out with the practical work, to Helen Irvine for help with identifying fumarate and to Jake Hodgson for his general enthusiasm for discussion. Unseen but not unappreciated were James Jameson's all night sampling efforts when he could have been enjoying his holidays.

I am extremely grateful for the financial support I have received from the BBSRC without which the project could not have taken place. Their generous provision has enabled me to live comfortably and has provided me access to first class biochemical engineering research facilities. The quality of these facilities is also in no small part due to the continued determination of the department and the academic staff in pursuing cutting edge research. Thanks are also due to Clive Osborne, Billy Doyle and Ian Buchanan who maintain these facilities.

On a more personal note thanks are due to my parents who brought me up to be interested in and to think about the world around me, who taught me to fear the Lord and spoke to me the words of eternal life. More recently they have had to put up with my absent presence as I spent several months with them writing at my desk in the attic.

Finally I would like to thank my fiancée Grace who will shortly become my helper, ally, encourager, confidant and companion and who has provided great motivation to finish writing quickly before our wedding.

Table of Contents

Table of Tables

Table of Figures

Table of Abbreviations

ACP	Acyl Carrier Protein
ACT	Actinorhodin
AG	ArabinoGalactan
AT	Acyltransferase
BBSRC	The Biotechnology and Biological Sciences Research Council
BDH	British Drug Houses
BSA	Bovine Serum Albumin
BST	Biochemical Systems Theory
CCCP	Carbonyl Cyanide-*m*-ChloroPhenylhydrazone
CER	Carbon dioxide Evolution Rate
CoA	Coenzyme A
DCW	Dry Cell Weight
DEBS	6-DeoxyErythronolide B Synthase
DH	DeHydratase
DMSO	DiMethylSulphOxide
DNS	DiNitroSalysilic acid reagent
ER	EnoylReductase
FBA	Flux Balance Analysis
FEBS	Federation of European Biochemistry Societies
FEMS	Federation of European Microbiology Societies
FOT	Flux Oriented Theory
GC	Guanine and Cytosine content of DNA
HEPES	N-2-HydroxyEhtylPiperazine-N'-2-EthaneSulfonic acid
HPLC	High Performance Liquid Chromatography
KEGG	Kyoto Encyclopaedia of Genes and Genomes
KR	β-KetoReductase
KS	β-Ketoacyl Synthase
MCA	Metabolic Control Analysis
mRNA	Messenger RNA
NMR	Nuclear Magnetic Resonance
NRRL	Northern Regional Research Laboratory
OD	Optical Density

ORF	Open Reading Frame
OUR	Oxygen Uptake Rate
PC	Personal Computer
PCR	Polymerase Chain Reaction
PHB	PolyHydroxyButyrate
PKS	Polyketide Synthase
PPP	Pentose Phosphate Pathway
PTS	Phosphoryl Transfer System
RED	Undecylprodigiosin
RMM	Relative Molecular Mass
RPM	Revolutions Per Minute
RQ	Respiratory Quotient
rRNA	Ribosomal RNA
RV	Red Variant
TCA	TriCarboxylic Acid
TE	ThioEsterase
THN	TetraHydroxyNaphthalene
TOC	Total Organic Carbon
tRNA	Transfer RNA
UCL	University College London
UV	Ultra Violate
VVM	Volume per Volume per Minute
WV	White Variant

The model used in this work can be found in appendix 1. The abreviations used in the model can be found in the same appendix.

1 INTRODUCTION

Antibiotics have proved to be probably the most important product of the biotechnology industry in terms of their impact on society. Their industrial production has led to the ability to effectively treat some of the more unpleasant diseases effecting mankind. One of the most challenging aspects of antibiotic production is making large enough quantities cheaply enough to make them affordable to all. In order to do this the titre of antibiotic produced by the organism of interest has to be increased. Many antibiotics are products of secondary metabolism. They are not used for growth and as a result are closely regulated and are often only produced naturally at low levels. This means that the organisms have to be modified to produce larger amounts. Unfortunately being products of secondary metabolism the pathways involved in their production are rather obscure and not generally well understood. For this reason strain improvement has had to take place via empirical and often random methods.

Recent years have seen the build-up of antibiotic resistance in pathogenic organisms. This has led to a renewed interest in finding new antibiotics which work by different mechanisms than those already known. One of the most important classes of compound used as antibiotics is the polyketides. It has been suggested that the polyketides that have been discovered so far display a 5 times higher incidence of biological activity than generally seen in other compounds (see Rohr, 2000). In the last 10 years there has also been a vast improvement in the knowledge we have of the way polyketides are produced and in the molecular tools to manipulate their production. It seems that it will soon be possible to rationally design polyketides and to combinatorally synthesise them (see section 1.3).

The prospects of this work coming to fruition are exciting. However the usual problems associated with achieving high level antibiotic production i.e. increasing titre will still have to be tackled. It has taken 50 years to improve penicillin production to the point where it can be classified as a bulk product. If possible it would be better to find a short cut to this process. The fundamental understanding of polyketide synthesis and the regulation of secondary metabolism should enable a rational approach to be taken to polyketide overproduction. At some point however if polyketide production is to be

increased beyond the traditional limitations, the rational design is going to have to take place at the level of the whole cell. This is where metabolic modelling and engineering can render assistance.

Metabolic modelling allows the metabolism of the cell to be investigated. The model can then be used to estimate the effect of making different specific changes to the cell's metabolism. The distribution of the flux through the different pathways resulting can be used to examine the control mechanisms involved in the metabolic network. This will allow more informed decisions regarding engineering metabolic changes.

The project aims to develop a metabolic modelling method known as flux balance analysis for studying secondary metabolite production in *S. erythraea.* Once the method has been suitably adapted it will be used to investigate the metabolism of *S. erythraea* in order to develop strategies for the overproduction of polyketides. This introduction will cover a number of areas key to the modelling of bacteria and to the overproduction of polyketides in *S. erythraea.* These will include a review of the taxonomy and biochemistry of the host, a review of the synthesis of polyketides and methods available to engineer this, a review of the methods available for modelling bacteria and the impact of post genomic techniques on metabolic engineering of these organisms.

1.1 System Under Investigation

1.1.1 Introduction

Saccharopolyspora erythraea was chosen as the organism to be investigated in these studies for a number of reasons. Firstly because the department has experience with it in growth and antibiotic production. Secondly through collaboration with Peter Leadlay's group in Cambridge we have access to strains producing novel polyketides and to overproducing strains. Thirdly the erythromycin biosynthesis pathway found in this organism is well characterised and all genes involved have been assigned functions

facilitating modelling. Fourthly the reasonably close relationship between this organism and *Streptomyces coelicolor* enables the use of the *S. coelicolor* genome for prediction of the pathways present in *S. erythraea*. The genome of *S. erythraea* has been almost completely sequenced, however gap closure is still underway and the sequence is not in the public domain (McKay 2001).

1.1.2 *Saccharopolyspora erythraea*

Description

Saccharopolyspora erythraea is an aerobic Gram-positive non-acid-fast actinomycete with a highly branched filamentous morphology unlike most of the members of its genus which grow in a fragmented manner (Zhou *et al.* 1998). It has been isolated from soil (Seno and Richardon 1986) although some members of its genus are associated with plant material (Goodfellow *et. al.* 1989). On agar growth is slow and colonies are formed which eventually differentiate producing aerial mycelia and spores in some strains. Associated with differentiation is the production of secondary metabolites. *S. erythraea* produces two known secondary metabolites: erythromycin which is well documented (McGuire *et. al.* 1952) and a red pigment of which less is known (Cortés *et. al.* 2002), both are polyketides. If the *streptomyces* genomes so far sequenced are anything to go by there may well be many more secondary metabolite gene clusters in *S. erythraea* (Bentley *et. al.* 2002, Ikeda *et. al.* 2003). Erythromycin is a macrolide antibiotic commercially produced and widely used in human medicine. The red pigment is composed of glycosylated and polymerised derivatives of flaviolin produced by a type III polyketide synthase. It is thought that the flaviolin moiety may undergo polymerisation to form melanin pigments with a role in UV protection (Cortés *et. al.* 2002). Growth in submerged culture displays an incomplete separation of growth and erythromycin or red pigment production. Strains producing the red pigment display lower erythromycin yields.

1.1.2.1 Taxonomy

The original isolation of *S. erythraea* was made between 1915 and 1920; sources vary on the actual year. It was originally named *Actinomyces erythreus* (Waksman 1919) and it is unclear when this became *Streptomyces erythreus*. The original isolate became inactive and although a number of strains claimed to be derived from that strain they were all found to be *Streptomyces griseus* (Labeda 1987). The international *streptomyces* project chose from two strains suggested by Waksman the one that was closest to his original description and made this the type strain. The name erythraea comes from applying Latin grammar to the Greek word erythros meaning red. All decisions regarding the taxonomy of *S. erythraea* mentioned so far were made using the classical taxonomic means of physiology and morphology. The first signs that *S. erythraea* did not sit happily in the *streptomyces* came from Kuznetsov *et al.* (1977). Kuznetsov studied the cell wall chemistry and found it to be type IV, which excluded it from the genus *streptomyces*. Instead it was placed in the genus *Nocardia*. This paper went largely unnoticed and the strain remained officially a *streptomyces* until Labeda rediscovered Kuznetsov's paper (Labeda 1987). He repeated Kuznetsov's work and extended it discovering that *S. erythraea* lacks nocardomycolic acids, excluding it from the genus *Nocardia*. Based on its morphological, chemical and physiological properties Labeda reclassified the strain into the genus *Saccharopolyspora* where it has remained ever since.

The genus *saccharopolyspora* became part of the family Pseudonocardiaceae in 1988 when Embley *et al.* (1988a) brought the mycolate-less wall chemotype IV actinomycetes into a separate family based on 16S ribosomal RNA considerations. This grouping has held up well under subsequent studies.

The type strain of *S. erythraea* is NRRL 2338 and was originally thought to come in 2 variants the red variant and the white (or grey) variant. This however has come into doubt as strains which behave as the red variant in one laboratory may behave as the white variant in another. In fact strains seem to revert phenotype with a rather high frequency. This was thought to be due to the rather high plasmid activity in the species (see Seno and Hutchinson 1986). However more recent work considers it likely to be a spontaneous pleiotropic mutation (Cortés *et al.* 2002) This instability has caused some difficulties in this work as will be recorded later.

It was hoped that the advent of 16S rRNA analysis would sort out a definitive classification of the actinomycetes. The method has been very useful for establishing or confirming deep divisions within the Gram-positive bacteria. However it has not proved so useful at working out the relationships between groups at the family level or even at the higher level of which group is the actinomycetes nearest neighbour (Embley *et al.* (1994), Ahmad *et al.* (1999), Ahmad *et al.* (2000)). Many phylogenetic trees have been drawn up based on 16 S rRNA analysis, trees can be found in Embley *et al.* (1988a), Embley *et al.* (1988b) Goodfellow *et al.* (1989), Bowen *et al.* (1989), Embley *et al.* (1994), Ahmad *et al.* (1999) and Ahmed *et al.* (2000). It is very difficult to draw anything conclusive from these trees. The family Pseudonocardiaceae usually seems to be grouped reasonably close to the *mycobacteria*; the papers seem less keen to give a degree of relation to the *streptomyces*. Work performed during this project seems to confirm a closer affinity to the *mycobacteria* than to the *streptomyces* (see section 2.9). Preben Krabben (a postdoctoral researcher at UCL) who has access to the *S. erythraea* genomic data considers *S. erythraea* to be closer to the *mycobacteria* and the corynebacteria than to the *streptomyces* (Preben Krabben personal communication).

To try to get round the problems associated with 16S rRNA methods Ahmad *et al.* (1999 and 2000) used sequences of genes encoding highly conserved proteins to create phylogenetic trees. These studies are aimed at the Gram-positive bacteria generally and so don't focus in on the actinomycetes as much as could be desired. They confirm the split between low GC and high GC groups. At a more detailed level they do not reach a consensus and sometimes produce highly questionable results.

To conclude, at the current time definitive assignment of the relationships between the families within the actinomycetes is not possible.

1.1.3 Polyketides

Polyketides are a large family of compounds that share a common underpinning biochemistry, which are produced in some bacteria, fungi and plants. These compounds display an enormous variety of structures and are often combined with other metabolites such as sugar units or amino acids (Katz 1997). The group is of interest industrially

because a large number of the polyketides found in nature display biological activity. Naturally isolated polyketides are now used as antibiotics, antitumour agents, immunosuppressants and veterinary agents (Lal *et al.* 2000). Recent advances in the understanding of the genetics and synthesis underpinning polyketide production has led to interest in manipulating these pathways to produce novel polyketides.

The biochemical similarity of polyketides comes from their synthesis as a chain of condensed acetate or propionate molecules (this will be discussed in more detail in section 1.2). This leads to a long carbon chain with keto groups at every other carbon hence the name polyketide. The structural and functional diversity of polyketides arises from the subsequent modification of the basic chain to yield structurally complex molecules.

S. erythraea produces erythromycin, a very important macrolide antibiotic and for this reason has undergone intensive research hence its use in these studies. It also produces a red pigment containing a polyketide moiety.

1.1.3.1 Erythromycin

Erythromycin is composed of three moieties: a 14 membered macrolactone ring and two deoxy sugars: desosamine and mycarose see Figure 1 (Katz 1997). The polyketide portion of erythromycin is synthesised from methylmalonyl-CoA and propionyl-CoA by a type I polyketide synthase. Erythromycin is most active against Gram-positive bacteria (Seno and Hutchinson 1986). It accumulates intracellularly and is active against the ribosome inhibiting protein synthesis and thus killing the cell.

1.1.3.2 The red pigment

The red pigment is composed of two moieties: a polyketide derived aromatic flaviolin moiety and a rhamnose sugar moiety see Figure 2 (Cortés *et. al.* 2002). The red variant produces high titres (1g/l) of red pigment under appropriate conditions (Ushio 2003). The red pigment also has antibiotic properties though these have not been exploited

commercially (Ushio 2003). The flaviolin component is, under certain conditions, thought to polymerise forming a brown melanin like compound which may have UV protective properties. This polymerisation means that the pigmentation seen may in practice be a mixture of compounds.

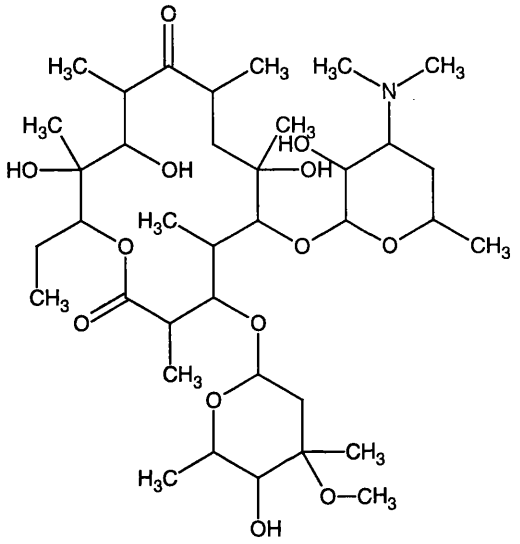
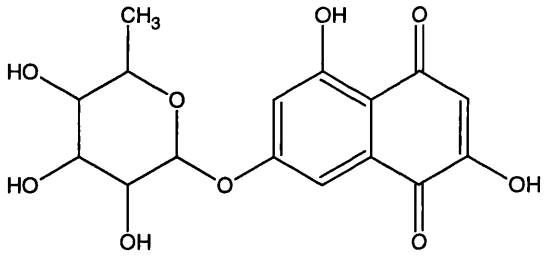


Figure 1: The structure of erythromycin A.



7-O-rhamnosyl flaviolin

Figure 2: The structure of the red pigment.

1.2 Synthesis of Erythromycin

1.2.1 Macrolactone Biosynthesis

1.2.1.1 Introduction

Polyketides are synthesised from acetate or propionate units in a manner similar to fatty acid biosynthesis. There are three main groups of pathways leading to the production of polyketides. The pathways are performed by large multifunctional enzymes called

polyketide synthases (PKS). Type I PKSs are modular in nature and types II and III are iterative in nature. Both types of enzyme contain many catalytic domains that sequentially perform additions and modifications of the acetate/propionate units involved (Shen 2003).

Iterative PKSs use some catalytic domains more than once in the production of the finished product. They tend to use only acetate building blocks and the keto groups are usually not reduced. They are often cyclised into aromatic rings systems. Important type II PKS products include actinorhodin and the tetracyclins (Lal *et al.* 2000). Important type III PKS products include tetrahydroxynaphthalene which is thought to form the basis of melanin like pigments in a number of actinomycetes (Cortés *et. al.* 2002).

Modular PKSs use each catalytic domain only once per molecule produced. They tend to perform more keto reductions than type II PKSs, with more flexibility in this reduction. Modular PKSs tend to produce large macrocyclic polyketides (Lal *et al.* 2000). The modular nature of type I polyketides has led to great interest in engineering these enzymes by domain swapping to produce novel compounds.

Until the recent work of Funa *et al.* (1999) type III polyketide synthases were only known from plants. However they have now been found in a wide range of Gram positive bacteria (Izumikawa et al. 2003). The main distinction in terms of mode of action of the type III synthase compared with types I and II is that CoA activated substrates are used directly by the enzyme rather than first attached to an acyl carrier protein (see below).

Although the type I, II and III classification system for polyketide synthases has proved useful the large number of PKSs coming to light in recent years has revealed quite a number which break the paradigm. For example iterative type I and modular type II PKSs. Shen (2003) hopes that these nonconformist PKSs will provide further diversity for engineering PKSs.

1.2.1.2 Synthesis of the Erythromycin Polyketide in *S. erythraea*

S. erythraea uses a type I polyketide synthase to synthesise erythromycin. The polyketide synthase is composed of three genes called *eryAI*, *eryAII* and *eryAIII*. The genes encode three proteins called DEBS 1, DEBS 2 and DEBS 3. Each protein contains two modules each of which can be considered as several domains. Each module catalyses a round of chain elongation, each domain catalyses one step in this process. Some of the domains are essential to chain elongation and some are nonessential. Nonessential domains perform alterations such as reductions to the chain as it is being produced (Katz 1997). The essential domains are

1. An acyltransferase (AT) domain
2. An acyl carrier protein (ACP) (for types I and II)
3. A β-ketoacyl ACP synthase (KS)
4. At the end of the PKS a thioesterase (TE) domain

The nonessential domains are

1. A β-ketoreductase (KR)
2. A dehydratase (DH)
3. An enoylreductase (ER)

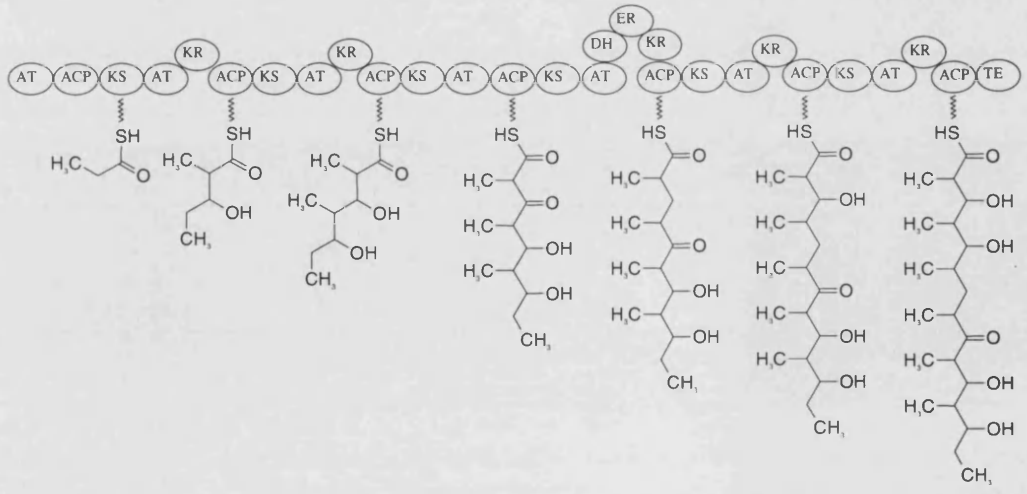


Figure 3: Schematic representation of the structure and operation of the erythromycin polyketide synthase. The polyketide chain is passed along the PKS from ACP to ACP elongating as it goes until it is released from the TE domain. This diagram is based on Lal *et al.* (2000)

Erythromycin synthesis starts with the synthesis of a polyketide chain by the PKS. A schematic overview of the process of chain elongation is presented as Figure 3. Polyketide synthesis starts with the attachment of a priming unit to the first AT domain on the PKS. The starting units used vary considerably between PKS systems but in erythromycin production the starting unit is propionate derived from propionyl-CoA. The AT unit controls the specificity of the starting unit. The propionate is then passed from the AT unit to the ACP unit where it forms a thioester with the ACP unit before being transferred to the KS domain. The nature of the extender unit used varies from system to system. In erythromycin the extender is provided by methylmalonyl-CoA. The methylmalonyl-CoA is recognised by the AT unit and attached via a thioester to the ACP. The KS domain then catalyses a Claisen condensation between the propionate and the methylmalonate. Decarboxylation of the methylmalonyl moiety provides the free energy for the condensation and for the breakage of the propionate ACP bond. The five-carbon chain then undergoes a reduction of the keto group derived from the propionate moiety to an alcohol group, this is catalysed by the KR domain. The five-carbon moiety is now ready to undergo a second round of chain extension in module 2. If both a KR and a DH domain are present then the keto group just created undergoes both reduction to a hydroxyl group and dehydration. If ER is also present then it undergoes enoylreduction forming a double bond in the chain.

The use of methylmalonyl-CoA as the extender unit means that the carbon is added in blocks of 3 carbons. Only two of these are used for the chain extension in each condensation reaction and the remaining carbon sticks out of the chain as a methyl group giving erythromycin its spiky appearance (see Figure 1). After 6 condensation reactions the polyketide chain reaches the end of the PKS. The TE domain then cleaves the thioester linkage to the final ACP and catalyses the cylisation of the polyketide into a macrolactone ring before releasing it from the PKS. Reviews of this whole process can be found by Katz (1997), Khosla (1997), Lal *et al.* (2000).

The molecules produced by this process tend to have a number of chiral centres and erythromycin is no exception. The stereochemistry of the chiral centres in macrolactones is not randomly assigned but follows rules discovered by Celmer (1965 a & b). The molecular basis of these rules has now been elucidated by Weissman *et al.* (1997). Chain extension utilises (2S)-methylmalonyl-CoA, the process inverts the

centre that undergoes condensation and the required stereospecificity is then introduced by epimerization if necessary.

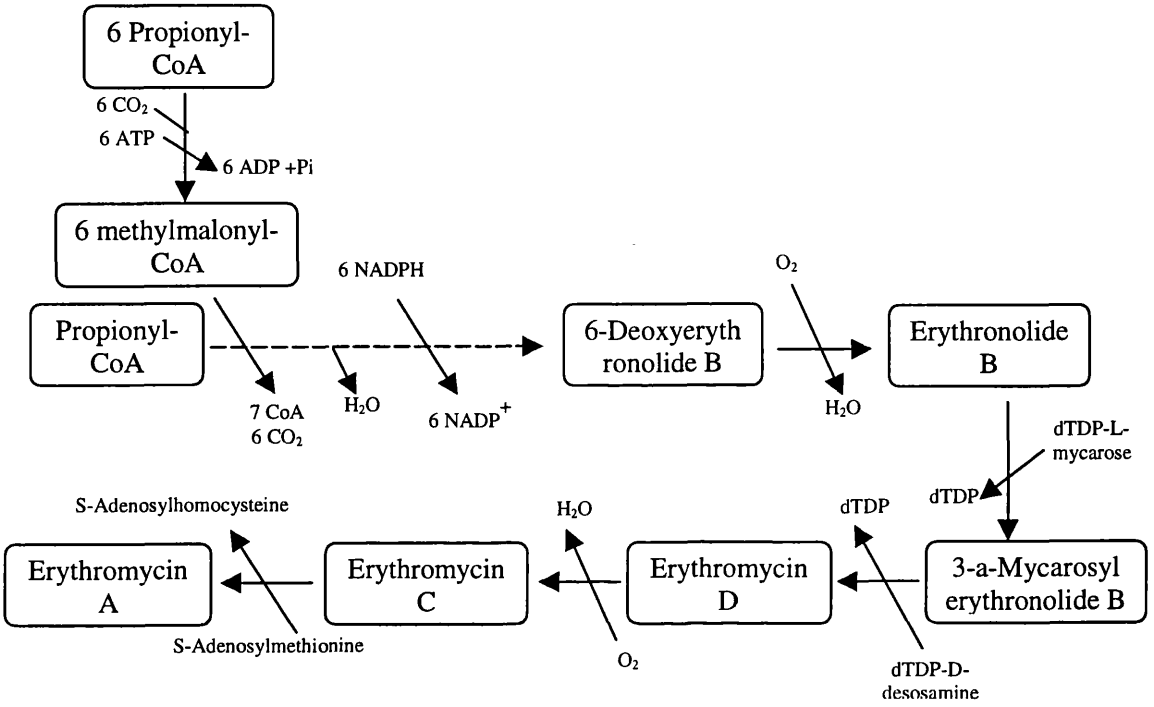


Figure 4: The biosynthetic pathway for erythromycin A. Synthesis of deoxysugars is shown in Figure 5

After the macrolactone ring is released from the PKS it usually undergoes a number of modifications. In the case of erythromycin these are in order: a hydroxylation at carbon 6, glycosylation with mycarose, glycosylation with desosamine, a hydroxylation at carbon 12 and an O-methylation on the mycarose moiety. Mycarose and desosamine are deoxysugars derived from glucose-1-phosphate (see section 1.2.2).

An overview of the reactions involved in erythromycin biosynthesis can be seen in Figure 4.

1.2.1.3 Relevance to Metabolic Engineering

The pathways of secondary metabolism are fed by the pathways of primary metabolism. If the analysis of Jeong *et al.* (2000) holds true for secondary metabolism these pathways should be much simpler and less interwoven than the pathways of primary metabolism. This should aid the metabolic engineer as less complexity is involved and there are less side branches to loose precursors to. However these pathways are not

adapted for carrying large fluxes and the cell's regulatory network is set up such as to restrict the flux through these pathways. Increasing flux through these pathways will probably not be straightforward.

There is evidence that increasing the level of expression of the genes of some secondary pathways does increase the level of flux through these pathways (Lomovskaya *et al.* 1999). This however will only hold true up to a point. Eventually the supply of precursors from primary metabolism will become limiting. The problem of manipulating the flux of these precursors will be much more complicated as they are often highly connected metabolites in an complex network of reactions. This makes it very important to be able to define how primary and secondary metabolism interface.

The main carbon provider for polyketide synthesis in *S. erythraea* is methylmalonyl-CoA, the sources of which in the cell are varied. The pathway also requires 6 ATP per molecule of erythromycin produced, 2 oxygen molecules and one methionine. The main interactions with primary metabolism are via propionyl-CoA or succinyl-CoA, ATP, TDP, NADPH and *S*-adenosylmethionine. The interactions and reactions involved in the deoxysugar biosynthesis will be considered in the next section. The overall reaction for production of the polyketide moiety of erythromycin is given below. This is based on the literature cited in the sections above and has been cross-checked by elemental and redox balances.

Propionyl-CoA + 6 methylmalonyl-CoA + 6 NADPH + 1 O_2 + dTDP-L-mycarose + dTDP-S-desosamine + SAM → Erythromycin A + 6 CO_2 6 $NADP^+$ + 3 H_2O + 2 TDP + SAH

1.2.2 Deoxysugar Biosynthesis

1.2.2.1 Relevance to Antibiotic Biosynthesis

It has been realised in recent years that the role of sugar moieties on antibiotics is not restricted to controlling pharmacokinetics such as absorption, distribution, metabolism and excretion (Kirschning *et al.* 1997). Rather the biological activity of many

antibiotics is now known to be dependant on the sugar moieties, on their type, number and attachment (Gaisser *et al.* 2000, Weymouth-Wilson 1997, Liu and Thorson 1994). It is thought that the sugars and the aglycone work together. The aglycone polyketide groups are often rigid providing a framework for the more flexible sugar groups. The deoxygenated carbohydrates are thought to convey biological activity by the flexibility of the glycosidic bonds and the balance of hydrophilic and hydrophobic domains (Kirschning *et al.* 1997).

The specifics of the role of deoxygenated oligosaccharides in antibiotics and antitumour agents that interact with DNA is starting to be unravelled (Kirschning *et al.* 1997). The role of deoxysugars in antibiotics targeting the ribosome such as erythromycin is however less clear (Kirschning *et al.* 1997, Weymouth-Wilson 1997). These deoxysugars are not only responsible for the advantageous activity of the antibiotics they are found in but also cause some of the side effects they exhibit (Weymouth-Wilson 1997). This has implications for improvement of existing antibiotics and design of novel antibiotics. Attempts to produce novel polyketide antibiotics will need to consider not only the polyketide chain but also the glycosylation of the chain (Zhao *et al.* 1998 b). How this is to be addressed has not been examined yet to the extent that the polyketide synthesis has been. The first signs of approaching this problem are starting to appear in the literature. The methods discussed are combinatorial in nature and are based on the flexibility of the glycosyltransferases (Gaisser *et al.* 2000, Kirschning *et al.* 1997). Enzymes involved in central metabolism are usually very specific for their substrate but those in secondary metabolism often display more flexibility. This flexibility allows for the addition of a number of different sugars using the same enzyme. If the enzyme could then be provided with a number of different sugars it might be possible to produce a polyketide with a range of glycosylations. Kirschning *et al.* (1997) have proposed a route that these might be introduced to the pathway without having to produce the nucleotide diphosphate versions of the sugar. This method uses a pathway for recycling deoxysugars which have become deactivated through hydrolysis of the high energy phosphate bond.

Another point to note with regard to the glycosylation of antibiotics is that in some organisms glycosylation is used as a method of inactivating the antibiotic whilst it is inside the producer thus protecting the producer from the poison it is making (Zhao *et*

al. 1998 a). Auto toxicity will have to be considered in the design of novel antibiotics and in the overproduction of current antibiotics. Inactivation by glycosylation may be an approach to this problem.

1.2.2.2 Biosynthetic Pathways

Erythromycin A, the main form of erythromycin produced in *S. erythraea* has two attached sugar moieties, L-mycarose and D-desosamine. These are derived from glucose metabolism and have to be synthesised especially for erythromycin production. The pathways for the production of these sugars are currently the topic of intensive research by a number of groups and although they seem to be reaching consensus there is not yet a definitive pathway. The pathway of Salah-Bey *et al.* (1998) has been used here however Trefzer *et al.* (1999) and Summers *et al.* (1997) have also presented pathways. Evidence for the Salah-Bey model is presented in Gaisser *et al.* (1997, 1998). Both sugars are synthesised from glucose-1-phosphate and their synthetic pathways are identical for the first few reactions. A summary of the pathways is shown in Figure 5.

These pathways have been derived from mutational experiments, by sequence comparison of the genes involved and by isotope incorporation experiments, firm conformation is still required.

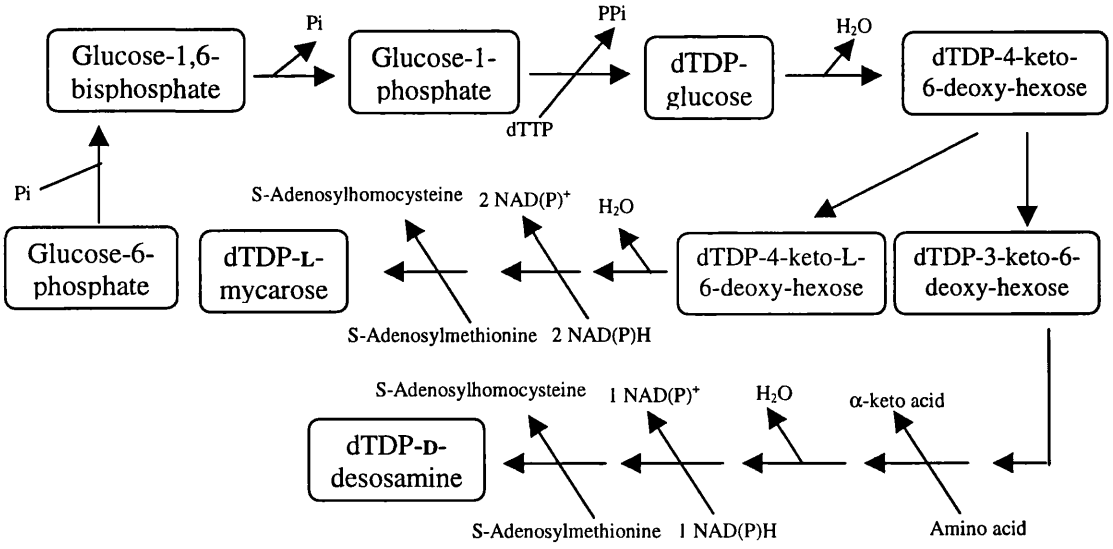


Figure 5: Pathways for the biosynthesis of deoxysugars.

Less is known at present about the transfer of saccharides to the aglycone portion of antibiotics. It need not, as was thought, occur after completion of the aglycone portion, in fact in erythromycin this is not the case (Kirschning *et al.* 1997). What is clear is that TDP hydrolysis provides the free energy for the glycosylation.

1.2.2.3 Relevance to Metabolic Engineering

The biosynthesis of deoxysugars is important, it is often the rate-limiting step in antibiotic biosynthesis (Gaisser *et al.* 1997). If the flux through polyketide production pathways is to be increased all pathways will have to be developed in parallel and this will include the glycosylation pathways. Fortunately all the deoxysugar biosynthesis pathways seem to have many common steps which should facilitate this.

In order to consider modelling such a pathway its interactions with central metabolism have to be established. The pathway interacts with central metabolism at a number of points. The precursor of the sugar comes from glycolysis in the form of glucose-6-phosphate, however the first steps in the pathway are shared in common with glycogen, trehalose and arabinogalactan metabolism up to glucose-1-phosphate. At this point there is a divergence of a number of pathways. The Glucose-1-phosphate molecules undergo nucleotide addition and the nucleotide involved changes with the metabolic fate of the sugar and the organism involved. Glycogen synthesis uses ATP (Parsons *et al.* 1978), arabinogalactan synthesis uses UDP glucose (Crick *et al.* 2001), 2,6-dideoxysugar synthesis uses TTP, 3,6-dideoxysugar synthesis uses CTP, lipopolysaccharide synthesis used GTP in *E coli* and CTP in *Yersinia enterocolitica* and UTP in deoxysugar biosynthesis in some *streptomyces* species (Trefzer *et al.* 1999). This then is a key branch point as it represents the point at which deoxysugar biosynthesis diverges from the rest of cellular metabolism.

The pathway also interacts with a number of cofactors from central metabolism (TTP, s-adenosylmethionine NADPH). When the sugars are attached to the erythromycin molecule the thymidine diphosphate is released. The pathways have a net requirement for reducing power. Three NADPH molecules are oxidised to produce the two sugar molecules used per erythromycin molecule synthesised. This brings the total NADPH

requirement for erythromycin synthesis to 9 moles NADPH per mole. If large yields of erythromycin are produced this will place a significant drain on the reduced cofactor pool of the cell.

The pathway interacts with a couple of amino acids. One of these amino acids is used in a transamination reaction, aminating one of the deoxysugars and producing an α-keto acid. This amino acid has not been identified in the literature. The other, methionine, is used in a number of methylation reactions. The side product of this is homocysteine. This either enters central metabolism via cysteine and propionyl-CoA (Voet & Voet 1990) or else is recycled back to methionine through methylation by methyl-tetrahydrofolate. One ATP is converted to adenine and ribulose-5-phosphate during the course of this reaction.

1.2.3 Synthesis of the Red Pigment

The structure of the red pigment has only recently been elucidated (Cortés *et. al.* 2002). and the details of its biosynthesis have not been investigated. However the flaviolin moiety is thought to be made in the same manner as in *Streptomyces griseus* (Funa *et al.* 1999). The overall reaction condenses five malonyl-CoA units into a chain which is then rearranged into tetrahydroxynaphthalene (THN). THN is spontaneously oxidised to flaviolin. Cortés *et. al.* (2002) believe that the rhamnose moiety is derived from TDP rhamnose which is known to be involved in cell wall biosynthesis. The overall reaction for synthesis of the flaviolin is. This is based on Cortés *et. al.* (2002) and on elemental and redox balances of the reaction.

$$5 \text{ malonyl-CoA} + 2\ O_2 \rightarrow \text{flaviolin} + 5CO_2 + H_2O$$

1.3 Metabolic Engineering of Polyketide Biosynthesis

1.3.1 Introduction

The modular domain based nature of the type I PKS makes it very amenable to manipulation. The theoretical basis for this is the flexibility of the modular domain structure of the PKS. If domains from different species can be mixed and matched then there is the potential to produce a huge variety of different compounds. As polyketides are such a rich source of biologically active compounds it has been reasoned that being able to produce such a huge variety should lead to the discovery of new biologically active compounds.

There are a number of types of domain in a polyketide synthase (see section 1.2.1.2) therefore there are a number of different ways in which a PKS can be manipulated to produce a novel compound.

1. Manipulating the TE domain. That is moving it so as to change the chain length of the polyketide.
2. Manipulating the first AT domain. That is changing the specificity of the priming unit.
3. Manipulating the subsequent AT domains. That is changing the extender units and thereby the side chains of the compound.
4. Manipulating the KR, DH and ER domains. That is changing the reduction level of the β-keto carbons.
5. Manipulation of stereochemistry. This is not so closely linked to a particular domain but can be manipulated nonetheless.
6. Post PKS modifications.
7. Manipulating the sugar groups. This is not controlled by the modular system but could prove very important.

For overviews of these methods see Lal *et al.* (2000), Katz (1997), Carreras *et al.* (1998).

1.3.2 Manipulating Chain Length

Manipulating the chain length of a polyketide will change the size of the macrolactone ring. It can be achieved by removing or adding whole modules to the PKS. This can be done simply at the genetic level as the genes are arranged linearly. There have been a number of examples of shortened chain length by removal of modules (Kao *et al.* 1994, Cortés *et al.* 1995). These are mostly proof of concept experiments and tools for further experiment rather than an attempt to deliberately engineer biologically active molecules. However (Ranganathan *et al.* 1999) demonstrated the use of this method, albeit in conjunction with method 3, to produce analogues of biologically active statins.

1.3.3 Changing the Priming Unit

This approach has been given a lot of attention as it is one of the simpler approaches to interspecies domain swaps. By adding an initial AT domain with a different specificity for the primer unit it is possible to change one of the side chains of the molecule. There has been great interest in primer domains with broad specificity. These allow new side groups to be added to the molecule which may add to its activity or provide a site for chemical modification. One example of a manipulation of this domain is the work of Pacey *et al.* (1998). They substituted the priming domain of the avermectin PKS into the erythromycin PKS in place of its own priming domain. The priming domain of the avermectin PKS is known for its broad specificity. Using this priming domain they were able to feed a variety of unusual fatty acids and show incorporation into the erythromycin molecule. One of the products had greater activity than normal erythromycin. By exploring the huge variety of organisms that display type I PKSs it should be possible to make use of a large number of starting units.

1.3.4 Changing the Extender Unit

The extender unit used in polyketide biosynthesis determines whether there will be a side carbon at a particular position and what the nature of that side group might be. It is

possible to control the side chains throughout the molecule by controlling the specificity of each domain. The first reported manipulation of extender unit was by Leadlay *et al.* (1993). Liu *et al.* (1997) used this method to rationally design a PKS to produce a specific compound. More recently Ranganathan *et al.* (1999) used this method to produce an analogue of an industrially important compound. If a number of domains with different extender unit specification can be found this could prove an important way of introducing change into macrolides.

1.3.5 Changing the Reduction Level

The level of reduction at each β-carbon in the molecule is determined by the presence or absence of the KR, ER and DH domains in the relevant module. By deleting or inserting these domains it is possible to manipulate the level of reduction of each β-carbon in the molecule. This has been demonstrated by McDaniel *et al.* (1997) who inserted an extra dehydratase domain from a different species and demonstrated that the product was reduced and dehydrated as predicted.

1.3.6 Manipulating the Stereochemistry

Type 1 PKSs produce chirally pure products. The manipulation of these properties is however not as simple as for the properties mentioned already. This is because there is no way of predicting what stereochemical properties the product of a PKS will have. Steps have been made towards understanding the basis of the chirallity of these compounds Weissman *et al.* (1997). Ranganathan *et al.* (1999) have demonstrated that it is possible to engineer the desired stereospecificity into a novel polyketide product. Being able to manipulate the chiral centres of potential drugs is very important as racemates can have very different biological activities.

1.3.7 Hybrid PKS and Nonribosomal Peptide Systems

Nonribosomal peptides are a group of secondary metabolites based on small peptide chains. The small peptide chains are synthesised in a manner very similar to the synthesis of polyketides. If the two systems could be fused it would lead to an even greater abundance of possible products. To this end Du *et al.* (2001) looked for natural occurrences of hybrid systems and found a number. Synthesis seems broadly to be divided between those where both parts are synthesised separately and ligated later and those where both parts are synthesised together. At the moment the fundamental knowledge is lacking for hybrid systems to be manipulated as PKS systems can. This may change however over the next few years as these systems are studied in more detail.

1.3.8 Post PKS Modifications

It is quite common for there to be modifications of the macrolactone ring after it is released from the PKS. Usually these are performed before any sugar addition but not always. In erythromycin there are two hydroxylations and a methylation. The enzymes performing these modifications are not coded for by the genes for the PKS and so have attracted less attention but they provide another level of combination to consider. Importantly they are able to perform modifications that are not possible within the PKS system. As yet there are no examples of gain-of-function modifications but there is an example of loss-of-function modification. Stassi *et al.* (1998) knocked out the two hydroxylation genes in the erythromycin pathway and produced a novel product 6,12-dideoxyerythromycin A.

1.3.9 Changing Deoxysugar Groups

Selection of deoxysugar moieties for attachment to macrolactone rings could be vital for producing novel biologically active compounds (Kirschning *et al.* 1997). Deoxysugar biosynthesis and glycosylation are not controlled by the modular system and are more

difficult to predict and engineer. For a novel sugar to be attached to a macrolactone the whole pathway for the production of the sugar might have to be introduced into the cell. This means a large number of genes must be introduced making the approach less attractive than manipulating the PKS. The enzymes responsible for attaching the deoxysugars are flexible enough in their specificity to attach a wide variety of sugars. This has led various people to consider combinatorial approaches *in vitro* (Gaisser *et al.* 2000, Kirschning *et al.* 1997, Zhao *et al.* 1998 b) but none has been demonstrated so far.

1.3.10 Combinatorial Biosynthesis

The polyketide synthases are receiving so much attention because in some ways they seem tailor-made for combinatorial rearrangements. The domain based structure makes manipulation relatively straightforward. By producing a library encoding many different domains it should be possible to recombine the domains to produce vast numbers of compounds that could then be screened for biological activity. Many have talked about combinatorial potential (Cane *et al.* 1998, Carreras *et al.* 1998, Lal *et al.* 2000, Marsden *et al.* 1998) and one or two have started to demonstrate the beginning of a combinatorial approach. McDaniel *et al.* (1999) generated a library of over 50 macrolides in a combinatorial fashion. Pacey *et al.* (1998) combined a changed priming domain with a library of potential starting units to produce 12 novel erythromycin derivatives. There has also been some interest in combinatorial approaches to glycosylation (see section 1.2.2.1).

1.3.11 Problems with Fusing Proteins

This technology is dependent on being able to splice part of one protein onto part of another protein. The splicing together of domains can work very successfully but can fail to work completely. For example Ranganathan *et al.* (1999) pointed out that out of a number of fusions they performed they obtained a variety of yields, they emphasised the need to define optimum fusion sites. Tang *et al.* (2000) looked at avoiding these

problems by combining entire protein subunits. This seemed to work well but cuts down on the flexibility of recombination at the domain level.

1.3.12 Interspecies PKS Transfer

In order to make the best use of the PKS system of synthesising polyketides it is important to isolate as many domains with distinct functions as possible. It has already been demonstrated that interspecies recombination works and a number of people have started looking at ways to combine domains from distantly related species. Bedford *et al.* (1995) have expressed a fungal polyketide in *Streptomyces coelicolor* and achieved product formation. Seow *et al.* (1997) cloned type II PKSs from soil bacteria that had not been isolated using PCR, they then expressed them in a *streptomyces* species making product. This is potentially important as it is thought that less than 1% of microbes can be cultured. More than 99% of the biodiversity of microbes is not accessible by normal cultivation. This provides a way of accessing "hidden" PKSs that exist in these organisms.

1.3.13 Conclusions

The relative simplicity of manipulation of PKS systems to produce novel polyketides has generated a lot of interest and research. The large number of manipulations possible combined with the modular nature of the synthases means that the potential for synthesis of new structures is almost limitless. As new products emerge from this field it will be essential to ensure that the capabilities exist to prepare the strains for industrial use. It will also be interesting to see whether the new compounds synthesised show the same high levels of biological activity associated with polyketides found in nature. It may be that nature has already selected and displayed the majority of structures with biological activity.

1.4 Metabolic Modelling

1.4.1 Introduction to Metabolic Modelling

Simple mathematical modelling of biological processes has been occurring for a long time. However in the last thirty years modelling strategies for biochemical processes have become much more complex. This complexity has been made possible by the increasing amount of information available regarding the underpinning biochemistry of life. However just as important has been the development of mathematical tools to represent that biochemistry. Now that complete genome sequences of organisms are being announced regularly some feel a shift in the approach of the biological sciences is inevitable. Palsson (2000) expects a shift from the current paradigm of reducing biological systems to their components to a paradigm where these components are integrated revealing the properties of the system as a whole. To deal with the huge complexity of the biological systems involved mathematical modelling will be key. This section will look at some of the methods available to metabolic engineers and their relevance to secondary metabolism.

1.4.2 Kinetic Models

Kinetic models are an attempt to define, predict and ultimately explain the dynamic behaviour of cells. In the simplest form these are models of growth such as the Monod equation such models are purely descriptive and not based on any information about the organism involved. At the other end of the spectrum are exhaustive white box kinetic models which attempt to take into account the whole biochemistry of the cell. This is the search for the system parameters which control the flux through a given pathway under any given set of conditions. A number of methods for investigating this have been developed including Metabolic Control Analysis (MCA), Biochemical Systems theory (BST) and Flux-Orientated Theory (FOT). All these methods have been shown to arrive at similar concepts of flux control (Stephanopoulos *et al.* 1998). Although white box kinetic models have far greater scope for gaining understanding and predictive power of an organisms metabolism than the other methods considered here

they are not appropriate for this project. This is because of the amount of detailed information required about the kinetic properties of the enzymes in the organism. This information is not available for *S. erythraea.* The method is also better suited to short pathways rather than whole organisms due to the large computational requirements.

1.4.2.1 Unstructured kinetic models

These are so called black box models where no information about the metabolism of the cells being modelled is considered. The kinetics of growth and product formation are observed and are fitted to models which can be used for online prediction and process optimisation.

One such model was produced by Elibol *et al.* (1999) for actinorhodin production in *Streptomyces coelicolor.* This was based on the Luedeking-Piret equation for growth. This model uses observations of growth and product formation and allows growth associated and non-growth associated product formation to be modelled based on a Monod like kinetics. No consideration of the metabolic network or the enzyme kinetics within the cell are used in this unstructured model. This method is limited in what it can achieve by its simplistic approach to the cell. They found it good at modelling glucose uptake and oxygen utilisation but not so good at biomass and product estimation.

1.4.2.2 Structured kinetic models

Between the black box modelling approach and the white box approach of MCA there exists grey box approaches generally known as structured models. These models do not try to assimilate all that is known about the cell. Rather they take basic simple aspects of metabolism which are well known, dealing with a much higher level than the individual enzyme. The kinetics of these higher level components is investigated and a model built based on these. These models have the advantage of being much simpler to build and acquire data for and because they are small they can be used for online process control (King 1997).

King (1997) constructed such a model of *Streptomyces tendae.* The model compartmentalised the cells into DNA, RNA, protein, biomass elements. Metabolites were classified into nucleotides, deoxynucleotides and amino acids. The feeds considered were glucose, ammonia and phosphate. He also included three regulatory mechanisms, the stringent response, protein control of DNA synthesis and DNA control of DNA synthesis. This very simple model of metabolism is basically a model of the central dogma. Using it he was able to predict behaviour in fed-batch culture and to control a fed batch strategy which significantly improved the titre of nikkomycin. Moreover by normalising the fermentation time it was possible to apply the same model to five other *streptomyces* strains with only minor modifications (King and Büdenbender 1997).

Roubos (2002) constructed a more complex structured model for *Streptomyces clavuligerus.* His model consisted of a metabolic network and either a neural network or a fuzzy logic module. Kinetic parameters that can be measured during the fermentation are measured. The metabolic network is then used to reduce the number of unknown kinetic parameters by linking various inputs and outputs. The remaining unknown kinetic parameters are then estimated by the neural net or fuzzy logic module based on the data available. These modules have to be trained on experimental data prior to use. The metabolic network was devised around primary metabolism and the model did not work as well during secondary metabolism. This was found to be in part caused by cell lysis and degradation of the product clavulanic acid.

1.4.3 Flux balance Analysis

1.4.3.1 Introduction

Flux balance analysis (FBA) methods rely on the principle of conservation of mass. By knowing that at any branch point in a metabolic network the mass entering must equal the mass leaving, and by knowing the structure of the metabolic network in the organism of interest, it is possible to perform a mass balance over the cell. These methods can yield information on the biochemical possibilities of a cell and on the maximum achievable yields of a product. They can calculate flux through known

pathways, identify new pathways and establish which of a number of possible pathways is being used.

1.4.3.2 Description

In basic flux balance analysis the pathways within the cell are used to construct a stoichiometric matrix of the reactions and the metabolites involved. The rows of the matrix correspond to the metabolites and the columns correspond to the reactions taking place in the cell. The values in the matrix reflect the stoichiometry of the metabolites involved in the reaction, substrates have a negative value and products a positive value. The complete matrix contains all the information concerning the stoichiometry of the reactions being modelled. By assuming the cell is in steady state i.e. that there is no change in the metabolite pool sizes in the cell the fluxes become directly interconnected. The matrix can then be thought of as a system of linear equations that can be solved to find the flux of each reaction in the network. These systems are usually underdetermined and some fluxes need to be measured in order to limit the matrix and solve the system. Fluxes used for measurement are usually substrates of the cell, extracellular products or biomass components that can be analysed simply. The basic steady state mass balance equation is:

Equation 1: $\mathbf{S \bullet v = b}$

Where

S = The stoichiometric matrix

v = The vector of reaction fluxes

b = The net accumulation rate vector

The cell is growing and accumulating biomass, however the rate of accumulation of metabolic pools in the biomass compared with the rate of turnover of these pools is negligible. This allows cells to be treated as though they were at steady state. At steady state there is no intracellular accumulation of metabolites therefore

Equation 2: $\mathbf{b = S \bullet v = 0}$

The reactions are arranged so that those that have been measured are at one end and those which have not been measured are at the other, the matrix is then split into G_m containing the reactions whose fluxes are measured and G_c containing those reactions whose rates are unmeasured. Likewise two vectors are composed, v_m contains the measured fluxes and v_c contains the unmeasured fluxes. The unknown fluxes can then be evaluated using the following formula. (This derivation is given in more detail in section 2.8.1.)

Equation 3: $$v_c = -\left(G_c^T\right)^{-1} G_m^T v_m$$

Some reactions cause problems, for example if there are two routes to the same product the linear equations of the two routes may be linearly dependant making resolution of the fluxes using linear algebra impossible, in such cases reactions are often lumped. Another problem is that for most organisms a complete metabolic network of the cell is unavailable making stoichiometric matrix generation difficult. This is usually dealt with by simplification or assumption. For example the central metabolic pathways are often assumed to be the same in organisms where they have not been verified as in those where they have, this is usually a valid assumption. For pathways where there is more uncertainty simplifications may be used, for example the biomass composition of a cell may be used to construct imaginary pathways directly from the building blocks such as amino acids and acetyl-CoA to the biomass ignoring the details of biomass synthesis. The validity of this method is dependent on accuracy of the biomass composition data used and the implications of this will be discussed in a later section.

A number of further analyses based on FBA method can also be used. If more measurements are made than are needed to solved the system the system is said to be overdetermined and the extra measurements can be used to check the accuracy of the model. Daae (1999) and Stephanopoulos (1998) have devised systems to minimise the errors in such a model by using a least squares method to fit the model to all the measured fluxes. In partial least squares fitting a solution is found which minimises the

residual for the measurements. That is it minimises the difference between the measured fluxes and the calculated fluxes.

Once a working metabolic network has been produced the network can be used to investigate and maximise the flux of metabolites to a product of choice. The sensitivity of the whole network, or of particular product fluxes, to certain perturbations can be found and the theoretical conditions for maximal product flux established.

Flux balance analysis can be used to identify points in the network that are rigidly controlled. The network has many nodes but usually only a few are instrumental in directing the flux to product, it is these principle nodes which must be studied for their rigidity (see Stephanopoulos *et al.* 1991). These nodes can be identified without exhaustive kinetic experiments by determining which nodes cause a large change in product yield when the flux partitioning through them is changed. Once the principle nodes have been identified their rigidity can be tested by altering conditions so as to encourage flux partitioning towards product, the resulting change in the partitioning is used to classify the enzyme. Nodes where the flux freely partitions between each branch according to the conditions are described as flexible. Nodes where the branch is dominated by one very fast pathway which takes most of the flux are known as weakly rigid. Nodes where the partition ratio is closely regulated are called rigid nodes.

Identification of these nodes is important as it determines which approaches are likely to be successful in changing the yield of product. For flexible nodes increasing the amount of enzymes in the pathway of choice should change the flux, for weakly rigid nodes it will also probably be necessary to reduce flux through the dominating pathway. For rigid nodes the enzyme's regulatory system will have to be removed.

1.4.3.3 Strengths and Weaknesses of Flux Balance Analysis

Flux balance analysis has a number of advantages over kinetic models. It is relatively simple, requiring information wise only a stoichiometric reaction network for the cell and no enzyme kinetic data. The calculations are also simple, so large reaction networks can be evaluated on a PC. It can provide information on the maximum

possible yield of product from a reaction network. It can provide information about which pathways are operating under certain conditions.

Flux balance analysis also has a number of disadvantages, which are to do with its simplicity. Although it may reveal that a certain yield is theoretically possible it does not yield much information on the kinetics involved in achieving that yield, i.e. how to achieve that yield practically. Although FBA can identify retrospectively the flux distribution in the metabolic network under a set of conditions it cannot predict what the flux distribution might be if those conditions were changed.

One of the biggest weaknesses with FBA is that some parts of metabolism cannot be effectively explored with it. If two pathways can be run such that they cancel each other out it is impossible to calculate how much flux there is in the two pathways. These fluxes can only be measured. Unfortunately measuring fluxes within the cell is difficult see section 1.4.6. Most systems have some pathways that are opaque to FBA. A particular problem is caused by the pathways of anapleurosis which are linearly dependant and hence cannot be resolved. Transhydrogenase activity also makes it impossible to determine the fluxes through the pentose phosphate pathway and the TCA cycle.

1.4.3.4 Applications

Flux balance analysis has been widely applied to both prokaryotic and simple Eukaryotic organisms. In industrial microbes it has usually been applied to the overproduction of primary metabolites such as amino acids. Recently a number of examples of its application to the production of secondary metabolites and to *streptomyces* species have been published. These however are discussed in the context of the results found in this work in the relevant sections especially section 4.3.5.

1.4.4 Metabolic Network Analysis

1.4.4.1 Introduction

The number of completed genomes is increasing all the time, but this only raises the question how should we interpret them in terms of whole organisms. This has led to an increasing realisation that it is the metabolism of the cell that defines its phenotype. A number of databases have arisen that deal with genomic information in terms of biochemical pathways (see Kanehisa *et al.* 2000, Overbeek *et al.* 2000 and Karp *et al.* 1999 for three such databases). One long-term aim is to be able to predict the metabolic network of a cell from the genomic data alone (see Huynen *et al.* 2000). Taking this a step further it has been recognised that merely predicting the pathways will not be enough, it is necessary to predict what the cell is phenotypically capable of doing with these pathways.

In order to achieve this two similar techniques for analysing metabolic networks have been developed and are compared by (Schuster *et al.* 2000). These methods do not consider pathways in isolation but in the context of the whole cell. In this context what constitutes a true pathway may be quite different from the traditional biochemical pathways. The two methods are called convex analysis and elementary mode analysis. These approaches attempt to distil the metabolic capabilities of the cell into a minimal number of pathways each describing a distinct capability from which by linear combination any possible flux distribution in the cell can be produced.

The stoichiometry of the reactions in the metabolic network and the requirement for no accumulation of intermediates within the network places constraints on what flux distributions can be achieved in the metabolic network. So for example valid flux distributions must start with reactants entering the cell, finish with products leaving the cell and all the reactions in-between must be connected product to substrate, with all cofactors regenerated in the process. The stoichiometry of the metabolic network usually does not constrain the flux distributions possible to a single solution. Rather there are an infinite number potential flux distributions which satisfy the constraints. All the possible flux distributions that satisfy the constraints are represented in this subset of the total number of possible flux distributions. This subset can be understood

geometrically as a subspace within the larger space represented by the total number of possible flux distributions. This subspace is referred to as the void space of the stoichiometric matrix. The shape of the void space is a high dimensional polyhedron extending from the origin. The limits of the void space within the larger space can be described by a set of vectors that follow the vertices of the polyhedron. These vectors are said to span the void space because by linear combination of these vectors any point within the void space can be described.

Both elementary modes analysis and convex analysis work by describing the void space of the stoichiometric matrix via such a set of spanning vectors. These spanning vectors represent the fundermenal flux distributions that the metabolic network is capable of. From these fundermanal flux distributions any possible flux distribution can be found by combination.

The key strength of this approach is its ability to describe the entire capability of the cell in terms of a finite number of extreme pathways. However it alos has a number of weaknesses. This form of analysis has been designed to be based on genomic information. If this information is not available for the cell of interest this method can only be used by extrapolation from a related organism for which the genome is available. It is also difficult to construct a complete metabolic network from genomic information as anotation is usually incompete and even in anotated genes cofactor specificity is difficult to assign. The pathways found may tell the theoretical capabilities of the cell but without restricting this with regulatory information about the enzymes involved theoretical yields found may be hard to achieve. While it may show that a flux distribution is theoretically possible it does not give any information on how this flux distribution might be achieved.

As networks get bigger other problems arise. The computation time rises greatly with increasing network size as does the number of elementary modes arrived at. This causes problems for large networks because the pathways or modes cannot be calculated in a reasonable amount of time. If a solution can be found the problem of how to interpret it arises. A network of 110 reactions produced 43279 elementary modes (Stelling *et al.* 2002). They cannot all be studied in detail. Whilst finding the one with the highest yield of product on substrate is straight forward, trying to gain insight into

the properties of the network is not, especially when most of the modes utilise most of the pathways in the network.

1.4.5 Genomics and Post Genomics

For many years biological processes have been modelled using simple black box models. In recent years however these methods have been supplemented by a number of white box models. These models make use of as much biological data as is available. White box modelling has been aided by the simultaneous development of genomics and other post genomics techniques which make far more data available than ever before. This section will outline some of these techniques and their relevance to modelling secondary metabolism in *S. erythraea.*

Over the past 50 years the life sciences have developed the techniques needed to study life at its most fundamental levels, that of the genetic repository of the cell and its transcription and translation. This started with the genome projects which sought to sequence organism's entire genomes. Later DNA microarray technology made it possible to study the levels of most gene transcripts in an organism simultaneously and this information became known as the transcriptome. Interest also grew in two-dimensional gel electrophoresis of proteins isolated from whole organisms. This procedure was combined with new mass spectrometry and image analysis techniques to record the protein expression levels of large numbers of proteins simultaneously. This became known as proteomics.

These techniques give a good picture of the cells macromolecular machinery. However some argue that the most important level to consider is not these but the metabolism of the cell (Fell 2001). To study this the field of metabolomics has grown up. Both mathematical modelling and analytical techniques are being developed to investigate the metabolome. Systems theory is being used to cope with the complexity involved in trying to understand whole organisms. Some researchers such as Bailey (1999) and Palsson (2000) see these developments as marking a shift in biology from reductionism to integration. Biological study has been simplifying the living world down to its most fundamental level, the genes that control life. Now that this fundamental level has been

reached it is possible to start to start to study how the components of life interact to make whole living organisms. So rather than breaking down and simplifying systems in order to study them, progressively larger and more complex systems will be studied. These technologies are all starting to have a large impact on metabolic engineering.

1.4.5.1 Genomics

Genomics is starting to have a big impact on the investigation of actinomycetes. The genome of *S. erythraea* has been almost completely sequenced, however the sequence is not yet in the public domain (McKay *et. al.* 2001). A number of genomes relevant to this project have been published. Notably *Mycobacterium tuberculosis* (Cole *et. al.* 1998), *Corynebacterium glutamicum* (Kalinowski *et. al.* 2003) *Streptomyces coelicolor* (Bentley *et. al.* 2002) and *Streptomyces avermitilis* (Ikeda *et. al.* 2003). Recent research indicates that the *Mycobacterium* and the *corynebacterium* show more similarity with genus *saccharopolyspora* in terms of general physiology than the *streptomyces*. So for example the cell walls of the *mycobacteria* and *saccharopolyspora* both contain arabinogalactan a cell wall polysaccharide. However the *mycobacteria* and the corynebacteria are adapted very differently for their environments. So as an obligate parasite *M. tuberculosis* experiences a very different environment than *S. erythraea* a soil bacterium. *S. erythraea* has a much larger genome than *M. tuberculosis* (8.7 megabases in *S. coelicolor* and 4.4 in *M. tuberculosis* Bentley *et al.* (2002) and Cole *et al.* 1998) which presumably allows it to survive in the much more variable and competitive environment of the soil. In this respect it is much more like members of the genus *streptomyces*. The *streptomyces* genomes published are useful for comparison when considering the more specialised areas of metabolism displayed by these organisms. For example differentiation, and secondary metabolism are seen in both the genus *streptomyces* and the genus *saccharopolyspora* but are not seen in the genus *Mycobacterium.*

The sequences of *S. coelicolor* and *S. avermitilis* have provided some interesting insights into these organisms. Genes for most of the basic physiology of the organisms are found in a central core region of the linear chromosome. This central region shows conserved gene order when compared with the whole genomes of other actinomycetes

which have much smaller genomes such as *M. tuberculosis* and *Corynebacterium diphtheriae*. The outer regions tend to contain less essential genes such as those for antibiotic production. These genome projects have revealed some interesting properties of these organisms. It has been known for a long time that these organisms have complicated regulatory mechanisms for undergoing differentiation. The genome of *S. coelicolor* is now thought to have nearly 1000 regulatory genes. The genome has also revealed that the four known secondary metabolites were just the tip of the iceberg. There are 18 more gene clusters that seem to code for secondary metabolites and it is possible to predict some of the structures of these compounds (Bentley *et. al.* 2002, see also Omura *et. al.* 2001). This has led to the discovery of some of these compounds that had gone undetected previously (Barona-Gomez *et. al.* 2003). A clearer picture of what products organisms such as these do produce will help in mass balancing fermentations. Without a reasonably close mass balance techniques such as flux balance analysis are hampered.

1.4.5.2 Transcriptomics

Transcriptomics has the potential to provide information one step beyond that revealed in the genome. Whilst the genome may give a complete list of the genes in the cell it gives no clue as to which are active when and to what extent. Ultimately genes are expressed through the activity of the proteins encoded. However it is not currently possible to simultaneously measure the activity of many proteins *in vivo*. Even measuring the level of expression of a reasonable proportion of the proteins in the cell by proteomics is still beyond the technology available. Transcriptomics assumes that the mRNA levels give a valid picture of protein expression and that protein expression gives a valid picture of protein activity. Whilst both these assumptions are far from perfect (Ideker *et. al.* 2001) microarrays do produce a large amount of data that can be usefully analysed which could not be otherwise obtained. So transcriptomics gives a step in the direction of working from genotype to phenotype on the whole organism scale.

There are two areas where transcriptomics ought to be able to offer vital information to modellers of metabolism. Firstly in the identification of which enzymes are being

expressed at any given time. This is important because although the genome defines the metabolic capabilities of the cell, the cell is unlikely to make use of all these capabilities simultaneously.

Secondly in the identification and investigation of regulatory networks (Stafford and Stephanopoulos (2001). Organisms like *S. erythraea* which display complex differentiation have complex regulatory networks to control the differentiation. More than 12% of *S. coelicolor* genes are predicted to have regulatory roles (Bentley *et. al.* 2002). Transcriptomics should help identify the sequence of interactions involved in regulatory events. This information will facilitate the elucidation of the structure and mechanism of the regulatory network. This will provide the information required to model these networks and to predict how to manipulate them.

There are no microarrays available for *S. erythraea* at present. However microarray technology has been developed for *S. coelicolor*. It is not yet clear whether it will be possible to used a *S. coelicolor* microarray with *S. erythraea*. However the information becoming available about *S. coelicolor* is interesting because it is hoped that the regulatory systems will prove similar in *S. erythraea*. Huang *et. al.* (2001) made a start on elucidating regulation in growth phase and production phase in *S. coelicolor*. They compared mRNA levels of about 5000 genes at 10 points during batch culture. They have been able to show which genes are co-ordinately regulated. This has enabled them to establish the extent of some of the regulatory mechanisms already known. The results also imply that there may be interaction between different stationary phase regulatory mechanisms. The availability of *S. coelicolor* microarrays should greatly accelerate the work in this area over the next few years.

1.4.5.3 Proteomics

Proteomics provides information one step closer to the phenotype than the transcriptome. It is however limited because generally only a few hundred proteins can be quantified using current techniques. As with the transcriptome the proteome only indirectly supplies information about the activity of the proteins expressed. Protein

level is not necessarily proportional to protein activity as the activity of the protein may be regulated by other proteins or by small molecules.

Proteomic data can be used for metabolic modelling in similar ways to transcriptomic data. It can provide an incomplete picture of which genes and hence which pathways are active under given conditions. It can help in the elucidation of regulatory networks, taking this a couple of steps further than the transcriptome. In the proteome not only is transcriptional regulation accounted for but also translational regulation. On top of this certain post translational regulatory events can also be identified. Some proteins are activated or inactivated by covalent modification eg phosphorylation, this can be identified. Some proteins are activated when cleaved by another enzyme these events can be identified too. However kinetic regulation of activity by small molecules is not detectable.

Proteomics has been used to investigate post-translational modifications in *Streptomyces coelicolor* (Hesketh *et. al.* 2002). In total 770 gene products were identified representing ca 10% of the genome. Some types of protein however were represented better than others. For example no sigma factors were detected which is unfortunate as these are known to play an important role in regulatory networks in *S. coelicolor*. One in seven proteins detected had undergone post-translational modification however only about 9% of the modifications could be identified. It was possible to identify proteins which had undergone: proteolytic cleavage, N-acetylation and adenylylation. Interestingly the results also show isozyme expression patterns in glycolysis which could be key to controlling flux through this pathway under different conditions.

1.4.5.4 Metabolomics

The metabolome is not as clear-cut a concept as the transcriptome or the proteome. The central dogma, DNA makes RNA makes Protein is a process in which some of the genetic information encoded in the genome is maintained in the sequences of the products. The step from the protein content of the cell to what the cell does in terms of its metabolism is a different kind of step altogether. Some see the metabolome as being

the levels of all the metabolites present in the cell, analogous to the proteome and the transcriptome. However this is currently a concept of limited use when investigating how genotype leads to phenotype. A more useful definition of the metabolome contains information on the flux of these metabolites though the metabolic network of the organism. Some have termed this the fluxome.

The metabolome in the limited sense of the pool sizes of the metabolites present in the cell is being investigated using analytical techniques. This usually involves rapid quenching of metabolism, extraction of small molecules and quantification of these species (Hajjaj *et. al.* 1998, Gonzalez *et. al.* 1997). Quantification usually involves a separation step such as liquid chromatography connected to a mass spectrometer eg (Groussac *et. al.* 2000, Soga *et. al.* 2002).

The flux through pathways in the cell can not be measured directly. Some fluxes can be calculated though by measuring fluxes into and out of the cell and performing flux balance analysis. This has been covered in section 1.4.3. Some fluxes cannot be calculated using flux balance analysis due to the structure of the metabolic network. In some cases specifically labelled carbon sources can be used to determine these unknown fluxes (Stephanopoulos *et. al.* 1998 chapter 9).

Little work has been done on determining the metabolome in organisms closely related to *S. erythraea.* However some of the techniques involved have been used for kinetic studies for example (Mira de Orduña and Theobald 2000). In this study the dynamic changes in small molecule concentration when a steady state is perturbed were investigated. The work showed that the rapid sampling and extraction techniques are applicable to *Streptomyces coelicolor.*

1.4.6 Measuring Flux

1.4.6.1 Introduction

Some of the models discussed here need flux data to be provided. In order for the model to be reliable the measured fluxes must be accurate. Accurate analytical methods for metabolites in complex and often non-homogeneous solutions like fermentation broths are difficult and suitable methods must be found. Some of the models described above are quite underdetermined and may require the measurement of many fluxes. The easiest fluxes to measure are external fluxes as these components accumulate rather than being in steady state like the intracellular metabolites. They can be separated from the biomass thereby reducing much of the background interference. They are usually reasonably stable and not subject to the cell's metabolism in the same way as internal metabolites. Likewise biomass constituents accumulate slowly over time and are not subject to the same high rate of turnover and low concentrations as internal metabolite pools. This makes their measurement easier and if possible it is preferable to measure only external metabolites in order to solve flux balance models. Some of the problems associated with making good measurements in filamentous organisms are considered below. Methods for measurement of external metabolites and biomass components are discussed in section 2.6. There is very little work on measuring internal metabolites in this project though so a brief outline of the salient issues is given below.

1.4.6.2 Growth

The considerable metabolic flexibility of actinomyces which allows them to radically change their metabolism under different circumstances also causes some difficulties if they are to be grown consistently for physiological studies. Small changes in the conditions for growth can lead to important differences in the metabolism of the organisms. Liao *et al.* (1995) offered some tips for avoiding some of these problems. They suggested: the use of carbon and nitrogen sources that can be quickly metabolised, use of spores for inoculation, and avoidance of transfer from complex media into defined media. *S. erythraea* is also very sensitive to changes in oxygen availability and the incorrect use of shake flasks can lead to low oxygen availability. The use of identical flasks with identical volumes of identical inoculum is vital. Pelleting can also

be a problem for filamentous organisms because the inside of the pellet will experience different nutrient concentrations to the outside. It is possible for the inside to be starving whilst the outside is still growing. Davidson (1992) investigated this problem in *Streptomyces coelicolor.*

1.4.6.3 Sampling

Sampling is an important part of performing accurate analytical work. If the sample is not representative of the general metabolic situation in the bulk broth then the accuracy of the analytical method becomes irrelevant. Even if the analytical work is accurate the result does not reflect the true situation. This means that sampling can introduce systematic errors.

There are a number of ways sampling can be problematic. First of all the sample has to be representative of the cell suspension it was taken from. The broth in bioreactors is sometimes not uniform. Filamentous organisms can cause problems by clogging sampling ports removing biomass from the sample. These problems can be overcome by good laboratory techniques and will not be considered further here. Wall growth and pelleting cause further problems. Wall growth means that there is more biomass in the reactor than there appears to be from the concentration in the broth. The biomass on the walls may be growing under different conditions to that in the bulk media. Wall growth is difficult to prevent. Pelleting also causes problems because analytical methods often use small volumes of media. In order for the volume to be representative the media must be reasonably uniform, pellets lower the uniformity.

Another problem occurs because the biomass does not cease to be active when removed from the culture. For time point measurements the bioactivity of the biomass with respect to the analyte has to be stopped. For external metabolites this is relatively straightforward, the supernatant can be removed from the cells by centrifugation or filtration, this should be done quickly and at low temperature to minimise biological activity see Mousdale (1998). For internal metabolites the situation is more complicated and a number of methods exist.

A fairly simple approach to performing analysis on internal metabolites has been developed by Mousdale (1998). He provides protocols for a large number of assays relevant to flux balancing which are based on simple laboratory equipment and are robust to the kinds of interference caused by biological samples. The approach used is to chill the cells, centrifuge them, wash the pellet and then inactivate the biomass with perchloric acid. This method uses a hazardous reagent (perchloric acid) and is chemically aggressive but can be used with stable metabolites for determining cell composition.

Another type of method which is much quicker in its inactivation is rapid chilling to extremely low temperatures. Two methods are reported in the literature, both involve immediate transfer of the sample into cold liquids; one uses liquid nitrogen (-80°C) and the other dry ice methanol (-40°C). Both methods were found to be equally effective by Hajjaj *et al.* (1998). These methods are very quick and not chemically aggressive but they do not remove the biomass from the broth before inactivation. Their main advantage is that they are quick enough to prevent rapid changes in most internal metabolites from effecting the results.

Recently non-invasive analytical methods for internal metabolites have been devised and these where applicable provide an important step forward as they represent the true situation in the cell. One of these (NMR based) will be discussed in section 1.4.6.5.

1.4.6.4 Extraction

Before samples can be analysed by invasive techniques the metabolites have to be extracted from the biomass. Three methods were recently reviewed by Hajjaj *et al.* (1998). They looked at acid extraction, alkali extraction and boiling ethanol extraction.

For the acid and alkali extractions the cells were quenched in methanol (-40°C) and permeablised by a chloroform/imidazole solution. Acid and alkali extractions were then performed by methods incompletely described. A neutral extraction method was used by (Mira de Orduña *et al.* 2000) for measuring intracellular glucose 6-phosphate in *Streptomyces coelicolor.*

In the boiling ethanol method the cells recovered from methanol quenching were extracted in boiling ethanol buffered with 10 mM HEPES at 80°C for 5 min. The samples were then cooled and concentrated, resuspended in water, centrifuged and analysed. More detail on this method can be found in Gonzalez *et al.* (1997)

The acid and alkali extractions are both chemically harsh and also selective, they are useful for some metabolites but can not be used for others. The boiling ethanol method was in contrast found to be much more effective. It recovered more of the metabolites, the metabolites were more stable in the ethanol and the possibility of a concentration step enabled lower levels to be analysed.

1.4.6.5 Analytical Methods

Analytical methods used in this area include: chemical tests, enzymatic tests, chromatography and NMR. Chemical and enzymatic tests are relatively simple to use and require little special equipment, however they are specific therefore if many metabolites are to be determined many tests are required. These methods are also susceptible to interference from the matrix so tests have to be chosen carefully (see Mousdale 1998). Chromatographic methods are more complex and require the correct equipment but they can analyse many components at once. Groussac *et al.* (2000) used two ion exchange chromatography procedures to analyse thirty intracellular metabolites of *Saccharomyces cerevisiae.* Soga *et al.* (2002) used capillary electrophoresis electrospray ionisation mass spectrometry to quantify 27 anionic metabolites from *Bacillus subtilis.*

Nuclear magnetic resonance (NMR) procedures are rather different. Isotopically labelled compounds can be identified and quantified within the living cell by NMR. This means that labelled metabolites can be added to cells and their rate of use and products evaluated. This can be used to determine the relative flux split at junctions in pathways. This means that the flux through an unknown pathway can be found by comparison with a known pathway it diverges from. Fluxes can also be found by adding a quantity of metabolite and measuring its disappearance relative to the size of the natural pool of that metabolite in the organism. These methods can be used to give

extra information that may be needed to make a stoichiometric matrix determined. The obvious suitability of this method for use with flux balance analysis has led to its development. For a summary of its use in FBA see Christensen *et al.* (1999) and Stephanopoulos *et al.* (1998). Simpler NMR techniques were used by Inbar *et al.* (1991) to find relevant information for secondary metabolite synthesis.

1.4.7 Conclusions

All the metabolic modelling methods discussed in this section have their own strengths and weaknesses. Flux balance analysis is conceptually simple and has been the most widely applied because the data required for its use is relatively straightforward to collect. The complexity and data requirements of metabolic control analysis has limited its use. Structured models are much simpler than kinetic models to use but have not attracted such wide spread attention. Pathway analysis is rather different from the kinetic and flux balance models. It is more interested in the structure of metabolic networks and hence its innate capabilities than in what it is doing and how. Although conceptually complex once a metabolic network has been constructed pathway analysis is the simplest to use as no measurments are required. It is also however the most limited in the information it yields. The interpertation of the data produced is also difficult because of the huge number of pathways found in large models. The choice of model will depend on the circumstances. In this work most modelling uses flux balance analysis although some elementary mode analysis has been performed on the network used.

1.5 Metabolic Engineering for Overproduction of Secondary Metabolites

1.5.1 Classical Production Improvement Methods

In the fifty years of commercial antibiotic production the titres of antibiotic achieved by the industrial processes involved have often risen dramatically. Erythromycin is one of the lower achievers in this respect with the titres of other secondary metabolites such as

penicillin being 5 to 10 fold higher (Minas *et al.* 1998). The types of improvement made in these fermentations fall into two general categories, improvement of the strain and improvement of the process.

1.5.1.1 Strain Improvements

Classical strain improvements have been achieved by a number of processes. The basic method is random mutagenesis followed by selection for improved performance. Some subtleties have been employed such as exposing strains to mutagenic substances whilst selecting for loss of regulation of tightly controlled enzymes by growth in the presence of an inhibitor analogue. Those organisms able to grow have overcome the inhibition and therefore the regulation. Recombination of strains is used to combine the best high producers and to cross back with the wild type to reduce deleterious mutations. Protoplast fusion is also used for similar reasons as recombination. These methods tend to lead to genetic instability of the strain. For a more complete description of this area see Lal *et al.* (1996)

Since the advent of genetic techniques for organisms such as the *streptomyces* a number of new approaches have become possible. For example the copy number of genes involved in antibiotic synthesis can be increased. One interesting approach using recombinant DNA technology was that of Minas *et al.* (1998). They introduced a gene encoding bacteria haemoglobin into an industrial producer strain of *S. erythraea* and significantly increased the titre of erythromycin by improving the oxygen uptake rate, the gene was also stable within the strain.

1.5.1.2 Process Improvements

Upstream process improvements come in two main forms, media optimisation and bioreactor improvement. Media optimisation has tended to be reasonably empirical in the past, however recently there have been attempts to make it systematic using multivariate statistics. One example is Ives *et al.* (1997) who used cluster analysis to identify which amino acids drew on the metabolite pool that clavulanic acid uses for its C_3 unit. These amino acids where then added to the defined medium used and an

increase in clavulanic acid production resulted. Another method developed by Gershater (2000) used principle component analysis to identify which combination of amino acids gave best product levels from a series of experiments on complex media.

Reactor design is dominated by the balance between oxygen transfer and shear damage. Antibiotic organisms are often filamentous, this makes them shear sensitive and limits the power that can be used to mix the fermenters and thus limits the oxygen transfer. Antibiotic production tends to require a lot of oxygen and so many commercial antibiotic processes are oxygen limited. In recent years innovative design of impellers has helped increase oxygen transfer without increasing shear damage. Another approach has been to look at the type of fermentation used. Traditionally fed-batch cultures are used. Continuous cultures are not favoured because they place commercial strains, which tend to be unstable, under great selective pressure because nutrient levels are always low in a chemostat. The low substrate levels can also restrict antibiotic production. Bushell (1997) and Lynch (1995) developed cyclic fed batch fermentations for erythromycin. These allowed the late exponential phase to be mimicked, combining both growth and antibiotic production. Using cyclic fed batch cultures they obtain significant increases in erythromycin production.

1.5.2 Metabolic Engineering Approaches

In contrast to classical strain improvement techniques in recent years a number of new techniques based on better knowledge about the systems involved in polyketide production have emerged. They often use genetic engineering techniques although to date the use of metabolic modelling has not had a big impact. Most of the techniques rely on intuitive understanding of how the systems work and so have not addressed trying to control the entire network of the organism.

1.5.2.1 *In Vitro* vs. *In Vivo*

A number of experiments have been performed with purified PKSs *in vitro*. This is fine at the test tube scale but as the scale of production is increased the cost of CoA activated compounds becomes prohibitive. In order to overcome this Pohl *et al.* (1998) have

found a chemical route to an extender unit that can be used instead of methylmalonyl-CoA. This compound is much less expensive than methylmalonyl-CoA and may provide a way forward for *in vitro* production.

1.5.2.2 Expression in a Foreign Host

The genetics of the actinomycetes is not as well developed as the genetics of some other bacteria. It is also more difficult to grow these filamentous species industrially. For these reasons there has been some interest in transferring PKSs to other species particularly *E. coli*. For a review of this area see McDaniel *et al.* 2001.

Transferring polyketide production to *E. coli* is not as simple as it might be. *E. coli* does not pantetheinylate the ACP domains of the PKS leaving them inactive (Khosla 1997) and it does not produce methylmalonyl CoA and therefore is unable to extend some polyketide chains such as that of erythromycin (Pohl *et al.* 1998). The pantetheinylation problem has been over come by co-expressing a holo-ACP synthase from *Bacillus subtilis* (Kealey *et al.* 1998). The methylmalonyl-CoA problem has partially been overcome by production of a synthetic analogue that can be fed to *E. coli* (Phol *et al.* 1998). Pfeifer *et al.* (2001) overcame both to synthesis a triketide lactone, the aglycone portion of erythromycin and a hybrid polyketide-nonribosomal peptide. Pantetheinylation was over come as above, the gene was integrated into the genome so as to block propionate catabolism which helps channel propionyl-CoA to the PKS. Precursor supply was overcome by upregulating a gene for conversion of propionate to propionyl-CoA and cloning in a gene for carboxylation of propionyl-CoA to methylmalonyl-CoA. Propionate fed to the cells is then converted to the precursors required for polyketide synthesis. The relevant PKS genes were also cloned in. The organism produced the aglycone portion of erythromycin at reasonable levels. They went on to optimise the yield by introducing a thioesterase from *S. erythraea* and using a fed-batch fermentation (Pfeifer *et al.* 2002). Using these strategies they were able to produced almost 0.2g/l erythromycin. This group also looked at using *S. coelicolor* as a more natural host for polyketide production. As the model streptomcyete more and better molecular biology tools exist for it than for *S. erythraea*. To improve precursor availability genes for the uptake of methylmalonate and its attachment to coenzyme-A

were cloned in from *Rhizobium trifolii*. This improved the yield three fold however the same strategy did not work in *E. coli* (see below).

Further work on optimising precursor supply in *E. coli* was performed by Murli *et al.* (2003). They improved on the plasmid systems for expressing the PKS genes increasing their stability in long fermentations. They also integrated the genes for precursor supply into the genome. As part of this work they tested three different routes for supplying the methylmalonyl-CoA. One pathway involved methylmalonate being supplied in the media and being converted by the cell to methylmalonyl-CoA. This led to 90% of the acyl-CoA metabolite pool in the cell being methylmalonyl-CoA but did not lead to polyketide synthesis. The second route was from succinyl-CoA to methylmalonyl-CoA. The third route was via carboxylation of propionyl-CoA. These two routes achieved methylmalonyl-CoA levels of about 30% of the acyl-CoA metabolite pool size. The level of polyketide production was however 8 times higher in the strain using the carboxylation route. Furthermore when both routes were introduced polyketide was preferentially produced from methylmalonyl-CoA made via the carboxylation route. The authors were not able to fully explain these results which clearly point to something unusual going on in the precursor supply for these reactions.

Another host changing strategy has been investigated by Rodriguez *et al.* (2003). Their underlying strategy is to use host strains that have already been selected for polyketide overproduction to produce newly discovered polyketides. They tested this by transferring the genes for the erythromycin PKS from a wild type organism to an industrial overproducer strain of *S. erythraea*. They found that the changes wrought in the overproducing strain were not in the PKS but rather in the promoter for the PKS and in precursor metabolism. They then went on to show improved titre for a novel polyketide based on the erythromycin PKS.

1.5.2.3 Inverse Metabolic Engineering

Inverse metabolic engineering or reverse metabolic engineering as it is sometimes known is a concept arrived at by Bailey *et al.* (1996). Rather than devise a strategy to achieve a metabolic goal from scratch the engineer looks for an organism displaying the desired phenotype. Once found the mechanisms underpinning it are studied and then

applied to the organism of interest. The method is of course limited by being able to find an organism with a suitable mechanism and on its compatibility with the host.

The most striking example of this strategy was that of Khosla and Bailey (1988) which was then followed up in many organisms including *S. erythraea*. They noticed that an organism which grew in an oxygen deprived environment produced haemoglobin under conditions of oxygen limitation. Knowing that oxygen limitation was a common problem in industrial scale fermentation they decided to clone the gene into *E. coli* to see whether it grew better under oxygen limitation. Using this method *E. coli* biomass could be more than doubled under conditions of oxygen limitation (see Bailey *et al.* 1996). The same strategy has been applied since to at least twelve different organisms achieving significant improvements in the productivity in each case. Of particular interest to this project is the work of Minas *et al.* (1998). They cloned the haemoglobin gene into the chromosome of *S. erythraea* and nearly doubled the yield of erythromycin from a commercial strain.

1.5.2.4 Unnatural Selection

One strategy that seems to be very powerful for changing the metabolism of organisms is to arrange the metabolism such that achieving the desired goal is in the best interest of the microorganism. If this can be achieved then the desired trait can be selected for. Whilst this is seen at a simple level with the use of antibiotic resistance to gain plasmid stability it is also possible to design subtle strategies to select for unnatural modes of metabolism.

Overkamp *et al.* (2002) give a good example of this. They knew that *Saccharomyces cerevisiae* was unable to grow without triose phosphate isomerase (TPI1) despite having all the enzymes necessary to grow by conversion of glucose to glycerol. They wanted to produce high yields of glycerol so they investigated what could be causing the problem. They identified two possible problems. Firstly the glycerol phosphate shuttle might be driving the pathway in the wrong direction especially if NADH was in short supply. Secondly the first step in the pathway required NADH which given a number of cytosolic dehydrogenases might be in low supply in the cytosol where glycolysis

takes place. These two mechanisms would lead to build-up of dihydroxyacetone phosphate (DHAP) the breakdown product of which is toxic to the cell. They knocked out both the dehydrogenases and the glycerol phosphate shuttle to open the way for balanced growth. Any cells which did grow however would have to convert the DHAP produced by glycolysis to glycerol to avoid autotoxicity. This put great selective pressure on the cells and viable cells were found and cultured. These cells however could not grow on the high glucose concentrations required for industrial production of glycerol. This was due to regulation of glycolysis which selects anaerobic growth for high glucose concentrations, however anaerobic growth is incompatible with production of glycerol as both branches of glycolysis then require NADH but only one is producing it. However by serial transfer to higher glucose concentrations they were able to select for mutants to overcome this regulation. The final strain grew on 400g/l glucose and produced 200g/l glycerol, the theoretical yield.

Another example of this is that of Flores *et al.* (1996). They wished to increase aromatic amino acid production which utilises phosphoenolpyruvate (PEP) as a substrate. PEP is used by the phosphotransferase system (PTS) in *E. coli* to phosphorylate glucose as it enters the cell. As the theoretical yield of PEP on glucose via glycolysis is 1 this means that very little PEP is available for overproduction of aromatic amino acids. To overcome this they knocked out the PTS. It was now in the interest of the cells to maximise growth using other systems to phosphorylate glucose and so they were able to select for growth on glucose. Strains were selected that could grow just as well on glucose as the wild type but had more PEP available to the cell.

Sauer (2001) has written at length about strategies to use selection to achieve improved strains of industrial organisms with a particular focus on use of chemostats as a highly selective environment. The combination of this kind of strategy with intuitive metabolic engineering could prove very powerful. What cannot be foreseen intuitively is often the regulation, but this is one aspect of metabolism that can often be overcome with selective pressure.

1.5.2.5 Directed Evolution and Genome Shuffling

Stemmer (1994) demonstrated that mutagenesis and recombination could yield improvement of the activity of individual enzymes with out knowledge of the catalytic mode of action. Since then there has been great interest in performing similar processes on whole organisms. This has become known as genome shuffling. Recently Zhang *et al.* (2002) performed genome shuffling on *Streptomyces fradiae.* This organism produced tylosin an antibiotic of similar structure to erythromycin. Starting from a wildtype organism they performed one round of mutagenesis and 11 strains displaying improved tylosin production were selected. Protoplast fusion was used to recombine these strains. The progeny were assayed and the seven highest producing strains selected. These underwent a second round of recombination. A further 7 strains with improved tylosin production were selected. The best two strains were then compared with a strain derived from the same parent that had undergone 20 years of classical strain improvement including 20 rounds of mutagenesis. The production was statistically indistinguishable from the classically improved strain and was roughly nine times that of the parental strain. The equivalent of 20 years work and 1,000,000 assays had been achieved in 1 year with 24,000 assays.

1.5.2.6 Overcoming Regulatory Structures

Regulatory mechanisms often prove problematic for metabolic engineering strategies. When faced with such regulatory issues it is often unclear how to approach the problem. One such problems is the flux through glycolysis in *E. coli.* Despite many attempts to increase the flux through glycolysis by increasing levels of the enzymes in glycolysis little progress was made. Koebmann *et al.* (2002) approached the problem from a different angle. Instead of trying to increase the supply of carbon to glycolysis they increased the demand for the products of glycolysis. They did this by introducing part of the ATP synthase which on its own converts ATP to ADP. This decreased the ATP levels and thus increased the demand for ATP and therefore the demand on glycolysis. By optimising ATP hydrolysis they were able to increase the flux through glycolysis 1.7 fold. They also calculated that under normal conditions nearly 100% of the control of glycolysis is due to demand i.e. due to something outside of glycolysis.

1.5.2.7 Over Expression of PKS Genes

Rodriguez *et al.* (2003) showed that most of the control of flux to polyketide production lies not within the PKS itself but in the expression of the PKS and in the supply of precursors to the PKS. A number of groups have tried to overexpress PKS systems with varying success. Rowe *et al.* (1998) placed the PKS genes for erythromycin synthesis under the control of the pathway activator gene *actII-ORF4*. And integrated them into the chromosome of *S. erythraea* red variant. The yield of erythromycin was increased 2.5 fold. However it should be noted that as the red variant has such a low production of erythromycin this is still less than the white variant wild type. Hu *et al.* (2003) tried a different approach. They developed a high copy number plasmid for *S. coelicolor.* Into this they placed the genes for the production of the aglycone moiety of erythromycin and achieved significant increase in titre over low copy number plasmids.

1.5.3 Considerations for Secondary Metabolism

1.5.3.1 Introduction

Much of the research performed on metabolism in bacteria has been investigating metabolism associated with growth. Yet many of the valuable compounds produced by bacteria are secondary metabolites. Growth of secondary metabolite producing bacteria is often considered in two phases: trophophase, associated with growth, and idiophase, associated with secondary metabolism, the two phases are not always completely separate. During secondary metabolism the primary and intermediary metabolic pathways drive secondary metabolism amoungst other things instead of growth. This requires large changes in operation of metabolism. Understanding these changes will be key for those wishing to successfully model and optimise secondary metabolism.

1.5.3.2 Primary/Secondary Interface

During the idiophase whilst product is being formed the primary metabolic pathways are present although they may be operating at somewhat different levels than those they operate at in trophophase. At this stage a large number of what might be termed

intermediary pathways may also be operating. Pathways for the recycling of the carbon in cellular components or storage compounds for example. Secondary metabolism often derives its precursors from these intermediate pathways. A clear understanding of the origin of the precursors of secondary metabolism is vital for metabolic modelling. Models based purely on primary metabolism are unlikely to yield useful information about the idiophase as they do not have the scope to deal with catabolism of biomass components.

1.5.3.2.1 Nutrient Limitation

Nutrient limitation often plays an important role in causing the move from primary metabolism to secondary metabolism in batch culture. McDermott *et al.* (1993) report that erythromycin production is seen under nitrogen, carbon and phosphate limitation in *S. erythraea.* They found that with carbon and phosphate limitation non-growth associated production was seen whilst with nitrate limitation apparently growth associated production was seen. Nutrient limitation has also been linked with other physiological phenomena in related species. Madden *et al.* (1996) reported that organic acid production (specifically pyruvate and α-ketoglutarate) was seen in nitrate limited *Streptomyces lividans.* Avignone Rossa *et al.* (2002) reported organic acid production in *S. lividans* under phosphate limitation. Production was also dependant on carbon source, there was little production with glucose but significant production with gluconate.

1.5.3.2.2 Organic Acid Secretion

Organic acid secretion has been observed in a number of *streptomyces* species. On agar the production of pyruvic acid and the subsequent fall in pH in unbuffered media has been linked to failure to form aerial mycelia and spores. However production in submerged culture is of more interest to this project.

Hockenhull *et al.* (1954) observed secretion of pyruvate and α-ketoglutarate at low levels in *S. griseus* grown on glucose. Feeding with the citric acid cycle intermediates fumarate and malate increased the levels of secretion of pyruvate and α-ketoglutarate. When cells were grown in the presence of arsenite which inhibits pyruvate and α-ketoglutarate dehydrogenases production of pyruvate was greatly increased. Transitory

secretion of lactate was also observed early in the growth phase, the lactate was reconsumed later in the fermentation. Growth under microaerobic conditions led to production of lactate which persisted for longer and reached higher levels until growth halted and cell lysis started to occur.

Doskočil *et al.* (1959) reported pyruvate production by *Streptomyces aureofaciens* in submerged culture. Pyruvate was produced during the early growth phase and was reconsumed later in the growth phase. Kannan and Rehacek (1970) measured organic acid production by starved *Streptomyces antibioticus* biomass on exposure to glucose. They found that succinic, citric, pyruvic, lactic and α-ketoglutaric acids were secreted.

Surowitz and Pfister (1985) investigated pyruvate secretion in *Streptomyces alboniger.* They found that pyruvate secretion was only seen with glucose, not with dextrin. This suggests a link to the rate of utilisation as glucose is a more readily utilisable substrate. Pyruvate secretion was reduced by the addition of adenine to the medium. The enzymes phosphofructokinase and pyruvate kinase were found to be around five times more active in cells grown on glucose than in cells grown on dextrin or glucose and adenine. Pyruvate decarboxylase and citrate synthase activities however were the same regardless of carbon source. In the glucose grown cells there was a mismatch in the activity between glycolysis and the TCA cycle with activity of glycolytic enzymes being around five times that of TCA cycle enzymes. This was concluded to be the cause of the secretion of pyruvate in *S. alboniger.*

1.5.3.2.3 Storage Compounds and Biomass Catabolism

Under conditions of nitrogen limitation *streptomyces* cells can accumulate carbon as storage compounds (see below). This means that accumulation of biomass can still be seen in the stationary phase. Whilst carbon storage compounds accumulate normal cellular components containing nitrogen such as protein and RNA show marked decreases. These storage compounds represent carbon that could be directed towards erythromycin biosynthesis.

Two kinds of carbon storage compound are common in nature, ones based on sugars such as starch and glycogen and ones based on neutral lipids. Both are found in the

bacterial kingdom and recent work has shown that both are found in *streptomyces* species (Olukoshi *et al.* 1994, Bruton *et al.* 1995) giving a good chance that they will be found in *Saccharopolyspora erythraea.* Another carbon storage compound that does not fall into either of these categories is sometimes found in bacteria. Some bacteria use polyhydroxybutyrate (PHB) as a carbon storage compound. PHB has also been found in *streptomyces* species (Ranade *et al.* 1993). Verma *et al.* (2002) investigated its occurrence in 12 *streptomyces* species and found it present in all at up to 11.8% of the dry cell weight. They also noticed a correlation between the consumption of PHB and the synthesis actinorhodin in *S. coelicolor.*

Trehalose is a storage compound based on glucose that is found in *streptomyces.* Karandikar *et al.* (1997) found that trehalose was accumulated during the primary growth phase and then was reduced to a lower steady level as the nitrogen source ran out. However Inbar *et al.* (1991) found that in *Streptomyces parvulus* its biosynthesis increased in the late log phase and by the end of actinomycin production trehalose was the major intracellular pool. Miguélez at al (1997) found that trehalose could be produced at any point in the life cycle and production was not restricted to a particular phase.

The papers of Olukoshi and Packter (1994) and Packter and Olukoshi (1995) reveal that triacylglycerols are used as storage compounds in a number of *streptomyces* species under nitrogen limitation. They estimate that triacylglycerol may make up to 80% of the cell volume in the stationary phase. The lipid was stored in membrane bound globular structures. A link to secondary metabolite production was suggested. Shim *et al.* (1997) studied the break down of triacylglycerides during stationary phase in *Streptomyces coelicolor* in order to demonstrate a link between this and polyketide production. They predicted that lipase activity should rise and isocitrate dehydrogenase activity fall as polyketide production started. They also predicted that neutral lipids would be utilised but not polar lipids. What they found correlated with this and they considered the storage lipids essential for polyketide production.

Ranade *et al.* (1993) found that *Streptomyces venezuelae* accumulated triacylglycerides, glycogen and polyhydroxybutyrate at various stages in its growth curve. Levels of

polyhydroxybutyrate were particularly high and were quickly consumed as the growth rate fell.

1.5.3.2.4 Precursor Biosynthesis

The rate of synthesis of erythromycin is dependent on the rate of synthesis of its precursors. It is therefore very important to understand this aspect of *S. erythraea*'s metabolism. A schematic representation of the way that various potential precursors of erythromycin enter polyketide synthesis is presented as Figure 6.

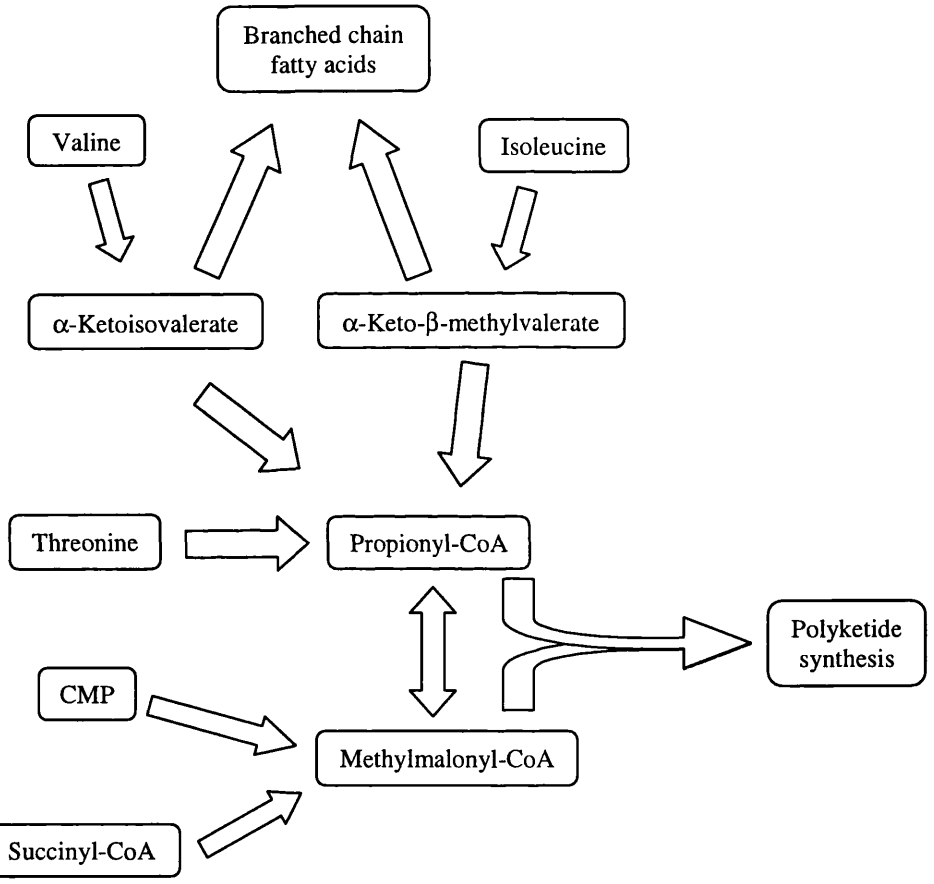


Figure 6: Precursor routes to polyketide synthesis.

One area that has been extensively investigated is amino acid catabolism, particularly catabolism of valine, leucine and isoleucine, which produce branched chain α-keto acids on transamination. Omura *et al.* (1983 and 1984) demonstrated that the catabolic products of a number of amino acids could be incorporated into tylosin a polyketide, closely related to erythromycin, produced in *Streptomyces fradiae*. The branched chain

organic acids produced by catabolism of valine and isoleucine are related in structure to methylmalonate and are a good potential source of precursors for polyketide synthesis.

Tang *et al.* (1994) investigated valine dehydrogenase in *Streptomyces ambofaciens* and *Streptomyces fradiae.* Knocking out valine dehydrogenase prevented polyketide production in the stationary phase. They suggested that it is not the only source of precursors for polyketide synthesis but that it is the main source employed when the carbon source runs out. Donadio *et al.* (1996) performed a conceptually similar experiment in *S. erythraea.* They knocked out propionyl-CoA carboxylase (an important step in the favoured route from valine and isoleucine to methylmalonyl-CoA) but found no impact on erythromycin production. They also pointed out that in *S. erythraea* the level of expression of this enzyme is in any case low and there are many routes to methylmalonyl-CoA and the actual route will depend on the conditions.

Smith *et al.* (1995) found that amino acid synthesis and catabolism were unregulated beyond simple feedback inhibition in *Streptomyces coelicolor.* This presumably gives it the metabolic flexibility to undergo different types of growth and metabolism as it situation changes. If the regulation amino acid catabolism is flexible this should facilitate metabolic engineering strategies for increasing the flux through these pathways.

Most routes to the methylmalonyl-CoA needed for erythromycin biosynthesis in *S. erythraea* go through either propionyl-CoA or succinyl-CoA. The route via propionyl-CoA has been discussed a little above. There have also been some papers investigating the route via succinyl-CoA. Bermúdez *et al.* (1998) investigated the activity of a number of enzymes in *S. erythraea.* These were isocitrate dehydrogenase, an enzyme in the TCA cycle, methylmalonyl-CoA mutase, leading from succinyl-CoA to methylmalonyl-CoA and Methylmalonyl-CoA decarboxylase, leading from methylmalonyl-CoA to propionyl-CoA. They found that isocitrate dehydrogenase and methylmalonyl-CoA mutase activities followed a growth associated pattern. Methylmalonyl-CoA decarboxylase did not show a clear change in activity as the cells entered stationary phase. Their results seem to suggest that the route via the TCA cycle is not very active during the stationary phase when most of the erythromycin production occurs. However it should be remembered that the actual fluxes of erythromycin are

very small compared to the fluxes which might be passing through these pathways during primary metabolism. It is thus not safe to assume this route has no role to play in erythromycin production especially as it is known to contribute to secondary metabolite production in other organisms.

Hsieh *et al.* (1994) investigated malonyl-CoA carboxylase, which catalyses the decarboxylation of malonyl-CoA to acetyl-CoA in *S. erythraea.* They found that its production corresponded to erythromycin production. They then knocked the gene out and found that erythromycin production was disrupted but could be restored by feeding propionate. Red pigment and sporulation were also disrupted by this knock out but were not restored by feeding propionate. They suggested that this might be due to the enzyme also catalysing decarboxylation of methylmalonyl-CoA producing propionate.

Murli *et al.* (2003) found some particularly interesting results relevant to the debate about whether succinyl-CoA or propionyl-CoA is the main route to methylmalonyl CoA. They introduced the genes for production of the aglycone moiety of erythromycin into *E. coli.* They also introduced genes for three routes to methylmalonyl-CoA. One via propionyl-CoA one via succinyl-CoA and one via feeding of methylmalonyl-CoA. All routes produced methylmalonyl-CoA in the cell but that produced by the propionyl-CoA carboxylase was used preferentially (see section 1.5.2.2). If this holds true in *S. erythraea* it could prove important for metabolic engineering overproduction. The mechanism of this preference should prove interesting if it can be elucidated.

From these findings it can be concluded that there are a large number of potential routes by which the precursors of erythromycin can be synthesised and they seem to be used to differing extents in differing circumstances. In order to be able to model secondary metabolism it will be necessary to have a more complete and flexible model than those generally used for flux balance analysis in order to reflect the complexity and flexibility of *S. erythraea.* The model will have to be able to deal with the breakdown of biomass for reuse. The number of routes by which this can happen is great therefore it will have to be larger and more complex than models for simpler organisms. It will also be necessary to have a much clearer picture of what is going on inside the biomass. It will not be sufficient to assume the biomass has the same proportional composition all through the process because this is clearly not the case in organisms that perform

secondary metabolism (see Karandikar *et al.* 1997 and King and Büdenbender 1997). To this end reliable methods of finding the composition of the biomass through the process will be essential.

1.5.3.2.5 Cofactor Regeneration

A number of cofactors are involved in the synthesis of erythromycin (see section 1.2) and of polyketides in general. The large number of reduction reactions that are often involved has led to interest in the supply of NADPH for these reactions. The production of NADPH by *streptomyces* strains and *S. erythraea* show some peculiarities as demonstrated by Roszkowski *et al.* (1971). Normally the pentose phosphate pathway produces two NADPH per glucose-6-phosphate entering. However in *S. erythraea* one NADPH and one NADH are produced. *S. erythraea* also uses an NADPH dependant isocitrate dehydrogenase and an NADPH dependent malic enzyme is also present. As will be demonstrated later (section 4.3.4.1) the regeneration of reduced cofactors has a huge influence on the flux through the pathways of central metabolism.

1.5.3.2.6 Regulation

Organisms that perform secondary metabolism are often thought of as undergoing a two-stage life cycle. The stages are called the trophophase and the idiophase. The traditional picture sees the depletion of the growth limiting substrate leading to the cell changing from trophophase to idiophase. So Karandikar *et al.* (1997) describe how depletion of nitrate leads to a phase change in *Streptomyces coelicolor.* This change was not simply between growth phase and stationary phase as the biomass continued to increase at reduced rate and DNA replication was also observed despite RNA and protein synthesis being halted. The simplicity of this picture of the transition from trophophase to idiophase has been questioned by Neumann *et al.* (1996). They found that during the exponential growth phase the cells at some point slow down and undergo a decision making process regarding how they will respond to nutrient limitation. Once the decision has been made to enter trophophase the cells are committed to it even if the nutrient levels are raised. The cells can not change the decision until they get into stationary phase. Which ever picture is true it is clear that such an intense upheaval of

the cells metabolism will be closely regulated and the understanding of this regulatory process is seen as key by many for the understanding of secondary metabolism.

The regulation of secondary metabolism has been the subject of much speculation for a long time. In recent years the picture of how it is regulated has started to come together from the wide array of different factors which were known to be involved somehow. There is still a long way to go in this respect but the next few years should see the unravelling of the control mechanisms in model organisms. This will be accelerated by the completion of a number of genomes and by the use of DNA microarray technology and proteomic technology (see section 1.4.5). However the genome has also pointed out how large the task will be. The *S. coelicolor* genome is predicted to contain 965 proteins with regulatory functions (Bentley 2002). Understanding and predicting the outcome of such large regulatory networks is going to require far more sophisticated approaches than are available presently. The wide array of factors are all interrelated therefore breaking them down is difficult.

Regulation of secondary metabolism is a very interesting area however it is not particularly well suited to being studied with flux balance analysis. Whilst the flux distribution in the cell is the result of regulatory interactions it does not reveal much about how the regulatory mechanisms act. The work presented here will not focus on this aspect of *S. erythraea*'s metabolism.

1.5.3.2.7 Auto Toxicity

One potential problem with producing antibacterial compounds in bacteria is that they may well kill the host. In nature this is often avoided by modifying the product to inactivate it within the cell. Most antibiotic producing organisms have more than one means of defence against their product (Zhao *et al.* 1998a). These methods include glycosylations and phosphorylations of essential components of the antibiotic. Alternatively resistance can be achieved by modifying the host machinery targeted by the antibiotic. For example erythromycin tolerance can be introduced by methylating the 23S rRNA (Champney et al. 2003). It may be necessary to introduce these systems into the host to protect it from the product. This is particularly relevant to systems expressed in species which do not naturally produce antibiotics.

1.5.3.3 Metabolic Considerations for *S. erythraea*

Most of the work considered in this section has been performed on *streptomyces* species. The genus *streptomyces* is reasonably closely related to the genus *saccharopolyspora* and we can expect their central and intermediary metabolism to be quite similar. The regulation of its metabolism however may not be so close. Superficially it appears similar with the idiophase triggered by the depletion of the limiting substrate and control of transcription plays an important part (see Reeve *et al.* 1998). However Bate *et al.* (1999) noted that the only *streptomyces* macrolide structural gene set completely sequenced has at least five potential regulatory genes whereas *S. erythraea*'s macrolide gene cluster contains no regulatory genes. In fact no regulatory genes effecting erythromycin synthesis have been found anywhere in the *S. erythraea* genome though this search has not been exhaustive. At the very least this implies that the elucidation of *S. erythraea*'s regulatory control of secondary metabolism is a long way behind the important *streptomyces* species. It could also mean that there are some fundamental differences between the two. This might fit taxonomically as *S. erythraea* seems to be more closely related to *C. glutamicum* and *M. tuberculosis* which do not display a *streptomyces*-like growth phase idiophase distinction. Moreover even the property of filamentous growth which seems to liken *S. erythraea* with the *streptomyces* is not displayed by most of the members of the genus *saccharopolyspora* (section 1.1.1).

1.5.3.4 Conclusions

Secondary metabolism is complex and is associated with a large number of issues which make both modelling and engineering of antibiotic production difficult. However it is this very complexity which requires the application of systematic tools for the study and manipulation of metabolism. It could be that this complex field that has frustrated traditional approaches to understanding and control is the perfect sounding ground for these new tools and approaches.

1.6 Project Aims and Objectives

Over the last 15 years there have been major advances in our knowledge of actinomycete biochemistry and in our understanding of the mechanisms of polyketide synthesis. Moreover techniques have been developed to manipulate the synthesis of polyketides. This makes the development of tools to model actinomycete metabolism vital. Such tools will allow the maximum potential of these organisms to be exploited. However in order to develop these tools some key issues have to be addressed. Some of these issues will be considered in this project. The findings will be applied to modelling of secondary metabolism in *S. erythraea*. The choice of modelling techniques is important for a project like this. The method has to be powerful enough to have the potential to give insight into the subtleties of secondary metabolism. It has to be able to take advantage of as much of the information that is available as possible. At the same time however it must be able to work without information which is not yet available for *S. erythraea*. Flux balance analysis was chosen because it uses information on the metabolic pathways within the cell. This information is reasonably available and becoming more so with the completion of more genomes. Flux balance analysis does not however require information about the kinetic parameters of the enzymes involved in these pathways. Very little of this sort of data is available. The limitations of flux balance analysis for use with organisms performing secondary metabolism will be discussed.

One of the key steps in this work will be going through the information available for *S. erythraea* and related organisms and using this information to construct a model that is able to be as accurate as possible to the known behaviour of the organism. The model constructed and the reasoning behind the choices made are presented in section 2.10. Accurate reproduction of the cells metabolic behaviour will also involve collecting information about the cell that is not available in the literature. For example the biomass composition of some actinomycetes changes during the stationary phase, however little quantitative data is available on these changes. In chapter 3 a near complete dynamic biomass composition for *S. erythraea* measured throughout a batch culture will be presented.

It has generally been assumed that biomass composition can be regarded as constant when performing flux balance analysis. However given the rather unusual metabolic behaviour associated with secondary metabolism this assumption needs to be tested before flux balance analysis can be applied to organisms like *S. erythraea.* The biomass composition information accumulated will be used to test whether using these measurements result in significantly better modelling of metabolism in *S. erythraea.* In chapter 3 a comparison of the results found by flux balance analysis both with and without consideration of biomass composition is presented.

The improved modelling method will then be used to explore the differences in metabolism between the exponential phase and the stationary phase. The results presented in chapter 4 give insight into some of the unusual behaviour exhibited by *S. erythraea* in batch culture. The comparison of differences in flux distributions taken at different stages of the fermentation will allow classification of key branch points of metabolism as to their regulatory flexibility. Comparison of actual flux distributions with theoretical distributions for maximal yield of erythromycin will be used to demonstrate how metabolism would have to be altered to allow such yields to be achieved.

Comparison of different strains of *S. erythraea* including an industrial strain will highlight significant differences in the metabolic properties of the organisms. These will be examined in the light of the relative production levels of polyketides exhibited by the organisms. For example in chapter 5 glucose uptake is shown to be much higher in the wildtype than in the industrial strain. This is shown to be potentially linked to a number of properties the strain displays that are conducive to overproduction of polyketides.

Modelling of secondary metabolism will enable the identification of key areas where there is potential for improvement of erythromycin production. It will also be used to highlight areas where our understanding is lacking. One such identified area is the production of organic acids. In chapter 6 a series of experiments are performed to try to gain further understanding of organic acid production in *S. erythraea.* The results are inconclusive and reveal that the complexities *S. erythraea*'s metabolism in this area run deeper still.

2 MATERIALS AND METHODS

2.1 Strains and Media

Strains of *S. erythraea* used in this work are listed in Table 1.

Table 1: Strains used in project

Strain	Description	Name used in text
NRRL 2338	White Variant wild type	White variant (WV)
NRRL 2338	Red Variant wild type	Red variant (RV)
CA340	Industrial overproducer	CA 340

Glycerol stocks of all these strains were produced in defined media and stored at –80°C.

Growth at all stages of fermentation was performed in defined media based on that of McDermott *et al.* (1993). Both a standard version without (carbon or nitrogen limitation) and a nitrogen limited version were used. Composition of the media (g/l in reverse osmosis water) can be found in Table 2 and Table 3. Components of the trace element solution were added in the order of the Table 3. After adding the magnesium sulphate the pH was adjusted to 2 with sulphuric acid. The rest of the components were then added. The solution was stored at 4°C. Glucose and salts were autoclaved separately, trace elements were added after sterilisation by sterile filtration, for fermentation, salts were sterilised *in situ*.

Table 2: Media composition

Component	Normal (g/l)	Nitrogen limited (g/l)
Glucose anhydrous	30	30
K_2HPO_4	7	7
KH_2PO_4	3	3
$NaNO_3$	11.1	2.38
Trace elements	10ml	10ml

Table 3: Trace element solution composition

Component	g/l
$MgSO4.7H_2O$	25
$FeSO4.7H_2O$	1.38
$CuCl_2.2H_2O$	0.067
$CoCl_2.6H_2O$	0.101
$CaCl_2.2H_2O$	1.38
$ZnCl_2$	1.04
$MnCl_2.4H_2O$	0.97
$Na_2MoO_4.2H_20$	0.035

2.2 Reagents and Chemicals

All reagents were purchased from BDH or Sigma and were of the highest quality available unless otherwise stated in the text.

2.3 Fermentation

Two bioreactor systems were used in this work. For fermentations described in chapter 3 and section 5.3 a 20 litre bioreactor was used (LH engineering Ltd 2000 series fermenter with 3000 series instrumentation and Adaptive Biosystems software control). The working volume was 13 litres. Agitation was set at 600 RPM using three Rushton turbine impellers. Sterilisation was performed *in situ* by moist heat at 121 °C, 1 bar pressure for 20 minutes.

For fermentations in chapters 4, 6 and section 5.2 a 7 L bioreactor (New Brunswick Scientific Co. Inc) with Bioflow 110 control units and Biocommand plus version 3.28 software control was used. The working volume was 5 litres and agitation was set at 700 RPM and achieved with two Rushton turbine impellers. Sterilisation was performed in an autoclave at 121 °C, 1 bar pressure for 20 minutes.

In all cases airflow was 1 VVM. pH control was at pH 7 and adjusted with 2 M phosphoric acid and 2 M NaOH. In early experiments antifoam control was applied using polypropylene glycol, however this was found to inhibit growth and all experiments reported here used Dow Corning "DB-110 A" silicon based antifoam. Off gas analysis was performed on a VG Gas Analysis Prima 600 mass spectrometer.

A two stage inoculum train was used for inoculation of the fermenters. All inoculum stages were performed in the unlimited defined media described above. A 500 ml baffled Erlenmeyer flask containing 100 ml media was inoculated from a cryogenically stored vial. This was grown for 48 or 72 hours depending on the strain and then split. Two 40 ml aliquots were used to inoculate two 2 L baffled Erlenmeyer flasks containing 400 ml media. 2 L flasks were grown for 48 hours. 5 L fermentations were inoculated with one such flask 13 L fermentations with 3 flasks to achieve an approximately 10% inoculation.

Shake flask experiments were performed in 2 L shake flasks. The inoculum was 40 ml taken from 100 ml grown in unlimited defined media in a 500 ml baffled Erlenmeyer flask.

2.4 Sampling

Samples were taken aseptically throughout the fermentation and placed immediately on ice. Dry cell weight measurements were taken without further treatment. Samples for intracellular biomass component analysis were spun down in a chilled centrifuge. The pellets were washed with ultra pure water before being frozen at –20 °C for later analysis. Samples for supernatant analysis were spun down in a chilled centrifuge. The supernatant was collected and frozen for later analysis.

2.5 Biomass Extraction

Biomass was extracted by the protocol of Mousdale (1998). By this method the contents of the biomass can be fractionated into four pools. 1) the soluble metabolite pool, 2) hydrolysed insoluble polysaccharides and hydrolysed nucleic acids, 3) the protein complement of the cell, 4) the lipid and cell wall material. By fractionating the biomass components like this it is possible to use quantification methods which would otherwise be rendered useless by the interference of other components.

Broth was collected as stated in the sampling section (section 2.4). 2 ml aliquots were spun down in a chilled centrifuge, washed twice with ultra pure water and frozen at –20 °C until use.

The pellet was made up to 2 ml with cold 0.2 M perchloric acid, resuspended and incubated at 4 °C for 16 hours. The suspension was then spun down again in a chilled centrifuge and the supernatant collected as the cold perchloric acid fraction. This was frozen at –20 °C until use. This fraction contains the soluble metabolite pool and was used for the trehalose analysis in section 2.6.7.

The pellet was made up to 2 ml with 0.5 M perchloric acid, resuspended and incubated for 2 hours at 70 °C. The suspension was then spun down at room temperature and the supernatant collected as the hot perchloric acid fraction. This fraction contains hydrolysed polysaccharides and nucleic acids. It was used for the DNA analysis in section 2.6.8 and the arabinogalactan analysis in section 2.6.6. Analysis of RNA from this fraction by the orcinol method as suggested by Mousdale (1998) proved impossible in *S. erythraea* see section 3.3.1.3.

The pellet was made up to 2 ml with 0.5 M NaOH, resuspended and incubated at 37 °C for 16 hours. The suspension was then spun down and the supernatant collected as the alkaline fraction. This fraction contains the cell's protein and was used for protein analysis in section 2.6.4.

The remaining pellet contains the lipids and the cell wall material. This was used for lipid analysis in section 2.6.9.

2.6 Analytical Determinations

2.6.1 Dry cell weight

Dry cell weight (DCW) was determined by the method of Davies *et al.* 2000. Duplicate samples of broth were used, 10 ml for early samples, 5 ml from the late exponential phase onwards. Samples were well mixed and then filtered through a Whatman type C glass fibre filter (Whatman PLC). The filter was washed 3 times with approximately 15 ml of reverse osmosis water. In all sections except chapter 4 and section 5.2 filters were dried before and after sample addition at 95 °C in a Mettler Toledo HG53 halogen moisture analyser. In chapter 4 and section 5.2 samples were

dried before and after addition of sample for 24 hours in an oven at 100 °C. Samples were weighted after each drying on an analytical balance and the DCW calculated.

2.6.2 Nitrate

In chapter 3 and section 5.3 nitrate was measured by the salicylic acid method of Cataldo *et al.* (1975). 20 µl of sample was added to 160 µl of 5% salicylic acid in conc.. H_2SO_4. This was thoroughly mixed and stored at room temperature for 20 minutes. 1.9ml of 2 M NaOH was added and the tube mixed and cooled to room temperature. The optical density of this solution was measured at 410 nm.

In chapter 4 and section 5.2 nitrate was measured by HPLC see section 2.6.15.2 on HPLC analysis.

2.6.3 Ammonium

Ammonium was measured by the indophenol blue method of Gerhardt *et al* (1981). 100 µl of sample was mixed with 1 ml of a solution of 10 g/l phenol and 5 mg/l nitrosopentacyanoferrate (III). 1 ml of a solution of 5 g/l NaOH, 8.4ml/l sodium hypochlorite solution (BDH GPR) was added and mixed. Samples were incubated at 37°C for 15 minutes. The optical density was then measured at 625 nm.

2.6.4 Protein

Protein was measured by the Bradford assay (Bradford 1976) using the method of Zor and Selinger (1996). Bradford assay reagent was prepared by dissolving 100 mg Coomassie brilliant blue G-250 (Sigma) in 50ml 95% ethanol. 100 ml of 85% phosphoric acid was added. This was made up to 1 litre with ultra pure water. It was stored in a dark bottle at 4 °C and mixed thoroughly before use. 400µl of sample was mixed with 800 µl of reagent. Proportions of sample and reagent could be changed to allow for different concentration ranges. Appropriate standards however have to be used. The optical density was measured at 595 nm and 466 nm with water as blank. OD_{595}/OD_{466} was used to calculate protein concentration from a standard curve. Extracellular protein was measured directly from filtered sample supernatant, intracellular protein was extracted first using the method of Mousdale (1998) see

section 2.5. Bovine serum albumin was then used as the standard for *S. erythraea* protein. Problems associated with the standard for this assay are discussed in section 3.3.1.6.

2.6.5 Glucose

In chapter 3 and section 5.3 glucose was determined by the dinitro salicylic acid (DNS) method based on Luchsinger and Cornesky (1962). DNS reagent was prepared by dissolving 16 g NaOH and 10 g of dinitro salicylic acid in 300 ml ultra pure water. Separately 300 g sodium potassium tartrate was dissolved in 300 ml ultra pure water with heating. Both solutions were mixed and made up to 1 litre with ultra pure water. 0.2 ml of sample was mixed with 0.2 ml of DNS reagent and incubated at 100°C for 5 minutes. Samples were cooled and 1.6 ml of ultra pure water was added. Optical density was measured at 540 nm.

In chapter 4 and section 5.2 glucose was determined by HPLC (see section 2.6.15.2).

2.6.6 Arabinogalactan

Arabinose and galactose were extracted from arabinogalactan in the biomass by the method of Mousdale (1998) (see section 2.5). The hot perchloric acid fraction was retained and neutralised with 0.5 M NaOH. The arabinose and galactose were then determined by HPLC according to the sugar protocol see section 2.6.15.2.

2.6.7 Trehalose

Trehalose was extracted from the biomass in the soluble metabolite pool by the method of Mousdale (1998) (see section 2.5). The cold perchloric acid fraction was retained. Trehalose was determined by HPLC according to the sugar protocol (see section 2.6.15.2).

2.6.8 DNA

DNA was measured by the Burton method (Mousdale 1998). Biomass was extracted by the method of Mousdale (1998) and the hot perchloric acid fraction retained. The Burton reagent was prepared as follows. 1.5 g of diphenylamine were dissolved in 100 ml 100% acetic acid, then 1.5 ml of conc.. H_2SO_4 was added. On the day of use 0.1 ml of 1.6% w/v acetaldehyde was added. 0.8 ml of reagent was added to 0.4 ml sample and mixed. Samples were incubated for 16 hours at 28 °C. Optical density was measured at 600 nm. Deoxyribose was used as standard and mass of DNA calculated from the average molecular weight per base for *S. erythraea* DNA (see section 3.3.2.1).

2.6.9 Lipid

Lipid was measured by the vanillin method (Mousedale 1998). Biomass was extracted by the method of Mousedale (1998) and the pellet from the alkaline fraction retained. The vanillin reagent was prepared by dissolving 200 mg of vanillin in 5 ml ethanol and made up to 100 ml with 85% ortho-phosphoric acid. 1.5 ml conc.. H_2SO_4 was added to the pellet and mixed. The sample was incubated at 100 °C for 10 minutes then cooled to room temperature. 75 µl of this was taken and 1.5 ml of vanillin reagent was added. The sample was incubated at room temperature for 30 minutes. The optical density was measured at 530 nm. Triolein (60% approx. sigma) was standardised with triolein (99% sigma) and used as standard. Problems associated with the choice of standard for this assay are discussed in section 3.3.2.5.

2.6.10 RNA

RNA was determined by the method of Benthin *et al* (1991). To a washed pellet from 2 ml of biomass 1.5 ml of 0.3 M KOH was added and the pellet resuspended. The sample was incubated at 37 °C for 1 hour with occasional mixing. The sample was neutralised with 0.15 ml 3 M perchloric acid, spun down and the supernatant collected. The pellet was washed twice with 0.5 M perchloric acid and the washings added to the collected supernatant. This was then made up to 10 ml with 0.5 M perchloric acid. The optical density was measured at 260 nm. The molar extinction coefficient for RNA is 10800 $M^{-1}cm^{-1}$ (Herbert *et al.* 1971). From this the molar concentration of RNA was calculated. The amount in g/l was calculated using the molecular mass per base of *S. erythraea* RNA calculated in section 3.3.2.2.

An alternative method based on detection of pentose sugars with orcinol was used in early experiments. Its use was discontinued because the presence of large amounts of the pentose sugar arabinose in the cell wall interfered with the assay. For this assay 3.0 g of orcinol were dissolved in 50 ml ethanol, 50 mg $FeCl_3.6H_2O$ were dissolved in 50 ml conc.. HCl. On the day of use the two solutions were mixed. 0.4 ml of sample and 0.8 ml of this reagent were mixed and incubated at 90 °C for 30 minutes. The optical density was read at 665 nm. With this method other pentose sugars in the cell such as deoxyribose and in this case arabinose have to be independently quantified and the appropriate subtractions made.

2.6.11 Glycogen

Glycogen was estimated by the method of Herbert *et al.* (1971). 5 ml of fresh cells was filtered onto a Whatman AF C filter and washed 3 times with 15 ml distilled water. Filter and cells were extracted for 3 hours at 100 °C in 30% KOH. 200 µl of this was added to 1.8 ml of 60% ethanol. The precipitate formed was spun down and washed with 60% ethanol. The supernatant was discarded and the pellet resuspended in 1 ml 0.2 M HCl. 0.2 ml of this suspension was taken and 0.2 ml of 5% (w/v) phenol added and mixed. 1 ml conc.. H_2SO_4 was added and mixed. Optical density was measured at 488 nm. Glucose was used as the standard.

2.6.12 Total Organic Carbon Analysis

Total organic carbon (TOC) analysis was performed on Shimazdo TOC 5050 analyser. Glucose solutions were used as standard. Samples were diluted 1:10 for analysis. Both supernatant and whole broth were analysed. Whole broth was sonicated prior to dilution to break up particles making the broth more uniform and avoiding clogging the sample needle. TOC data was used to check the carbon balances. This was done by comparing the total amount of carbon found in the broth by the analytical methods used with the amount of carbon found by the TOC analyser.

2.6.13 Erythromycin

Erythromycin was concentrated by the method of Heydarian *et al* (1998) and assayed by HPLC. 10 ml of broth supernatant was filtered through a 0.2 μm filter and applied to a C18 bond elute column (Varian inc.). The column had been previously prepared by wetting with 5 ml of 0.5 ml diethyl amine and washed with 5 ml 10 mM phosphate buffer solution. The erythromycin was bound to the column and the column was washed again with 5 ml 10 mM phosphate buffer. The erythromycin was then eluted in 1 ml methanol and placed in a HPLC vial. The vial was stored at –20°C until analysis. The erythromycin was then quantified by HPLC (see section 2.6.15.1).

2.6.14 Organic acid analysis

Pyruvate, fumarate and α-ketoglutarate were determined by HPLC analysis of undiluted filtered broth (see section 2.6.15.2).

2.6.15 High Performance Liquid Chromatography

HPLC was performed on two systems using identical procedures. A Beckman System Gold apparatus with UV and refractive index detection and a Dionex summit HPLC system with UV and refractive index detection.

2.6.15.1 Erythromycin

For erythromycin determination a PL 5u C18 ODS 15 cm column (Polymer Laboratories ltd) was used at 60 °C. The mobile phase was 55% 10 mM potassium phosphate (pH 7) and 45% acetonitrile pumped at 1ml/min. Injection volume was 20 μl. Detection was performed at 215 nm. Analysis time was 15 minutes. Quantification was performed using peak area based on standard solutions of erythromycin A.

2.6.15.2 Organic Acids and Sugars

For all other analytes the column used was an Aminex HPX-87H 30 cm at 60°C. If greater separation was required 2 of these columns were run in series. The mobile phase was 5 mM H_2SO_4 pumped at 0.6 ml/min. Organic acids were detected by UV

absorbance at 215 nm. Nitrate was detected by UV absorbance at 225 nm. Sugars were detected by refractive index. Quantification was performed using peak area based on standard solutions of the analytes. Injection volume varied by analysis. For Biomass sugars 50-100 µl were used depending on concentration. For analysing the supernatant of fermentation broth if nitrate was also to be measured a 7.5 µl injection volume was used. If nitrate was not to be measured a 20 µl injection volume was used. When intracellular metabolites were measured the injection volume was 100 µl.

The use of this column for detecting nitrate is a little unusual. Nitrate is not separated on this column but elutes with the solvent front. Normally this would not make for accurate analysis however the UV absorption of nitrate is so strong that the relatively high concentrations being monitored here have a far higher absorption than the rest of the solvent front allowing accurate determination of the nitrate concentration. This is convenient because with this one HPLC analysis it is possible to quantify the nitrogen source, the carbon source and the three organic acid products simultaneously in undiluted filtered broth.

2.7 Calculations

2.7.1 Calculation of Uptake and Secretion Rates

In batch cultures nutrient utilisation and product formation rates change during the course of the fermentation. Flux balance analysis requires that these rate be known. Rates were determined at selected points in the fermentation from the fermentation profiles. To do this curves were fitted to the profile using non-linear regression and rates calculated by differentiation. All rates were calculated as milimole per gram dry cell weight per hour. The growth curve was fitted to Monod kinetics and DCW values were interpolated from the equations derived. These interpolated DCW values allowed the calculation of the specific uptake rates at any point in the fermentation profile.

During the exponential phase Monod kinetics were assumed for all rates unless curve shape dictated otherwise. Monod type kinetics were linearised by plotting the natural logarithm of the variable against time. Some profiles which follow Monod kinetics are not as simple to fit as the growth curve. For example the glucose uptake rate is related

to the amount of biomass and so is related to Monod kinetics. The glucose concentration however falls instead of rising. In this situation which arose for both the glucose and the nitrate concentrations the following method was employed. The concentrations measured at the sampling points were subtracted from the initial concentration. The calculated values can then be treated as for the Monod kinetics. This method gives large errors for early results. This is because for the early results it involves finding a very small number by subtracting one large number from another both with associated errors. This acts to magnify the errors. For this reason data from the first 15 hours was not used in modelling. If problems remained after this time polynomial fitting was used to find these uptake rates.

In the stationary phase regression to linear or polynomial equations was performed using the standard regression tools in Excel. It was often necessary to break profiles down into sections which were fitted independently because of the complexity of the profiles.

2.7.2 Calculation of OUR and CER

The oxygen uptake rate (OUR) and carbon evolution rate (CER) were calculated as follows.

Equation 4:
$$gasflow_{in} = \frac{f \times 60 \times 1000}{V_r.mgv}$$

Equation 5:
$$gasflow_{out} = \frac{f \times 60 \times 1000}{V_r.mgv} \cdot \frac{N_{out}}{N_{in}}$$

Equation 6:
$$OUR = gasflow_{in}.O_{2in} - gasflow_{out}.O_{2out}$$

Equation 7:
$$CER = gasflow_{out}.CO_{2out} - gasflowin.CO_{2in}$$

Where
- $gasflow_{in}$ = air flow into fermenter (mmole/l/h)
- $gasflow_{out}$ = air flow out of fermenter (mmole/l/h)
- f = air flow into fermenter (l/min)
- V_r = Working volume of the reactor (l)
- mgv = The molar gas volume of an ideal gas (l)
- N_{2out} = Mole fraction of N_2 in gas out of the fermenter
- N_{2in} = Mole fraction of N_2 in air into fermenter
- O_{2out} = Mole fraction of O_2 in gas out of the fermenter
- O_{2in} = Mole fraction of O_2 in air into fermenter
- CO_{2out} = Mole fraction of CO_2 in gas out of the fermenter

CO_{2in} = Mole fraction of CO_2 in air into fermenter
OUR = Oxygen uptake rate (mmole/l/h)
CER = Carbon dioxide evolution rate (mmole/l/h)

Note the working volume of the reactor falls during the course of the fermentation due to sampling. This has to be accounted for in the calculations.

2.8 Modelling

Flux balance analysis is the main form of modelling used in this project. It makes use of linear algebra techniques. Modelling was performed in MatLab version 6.0.0.88 using Flux Analyzer (version 4.1), a metabolic modelling suit generously provided by Steffen Klamt (Klamt *et al* 2003). Steffen generously customised the software to provide χ^2 statistics for more degrees of freedom than in the original version. Flux distributions were calculated using the variance weighted least squares method assuming 1% variance on all measured or assumed fluxes. In this section the techniques used are described and explained. The description follows that given by Stephanopoulos *et al.* (1998).

2.8.1 Calculation of Flux Distributions

Flux balance analysis is based on the principle of conservation of mass through metabolic networks. The metabolic network devised for the cell is converted into a stoichiometric matrix. In this matrix the stoichiometric coefficient of each metabolite in each reaction in the cell is recorded. Each column represents one metabolite and each row represents one reaction. The value of each element in the matrix is the stoichiometric coefficient of that metabolite in that reaction. Metabolites consumed in a reaction are given a negative stoichiometric coefficient and metabolites produced in a reaction are given a positive coefficient. This matrix is denoted **G**.

The rate of accumulation of any intermediate in the matrix can be calculated by multiplying the rates of the reactions it is involved in by the stoichiometric coefficients for the reactions. When considering the rates of accumulation of all the intermediates this can be written in matrix notation as

Equation 8 $\mathbf{r} = \mathbf{G}^T\mathbf{v}$

Where **r** = The vector of the rates of accumulation of the intermediates in G
G = the stoichiometric matrix
v = The vector of the rates of the reactions in G
^{T} = Super script *T* indicates that the preceeding matrix is transposed

Two assumptions are made a) that the cell grows at steady state and b) that the rate of increase in metabolic pool sizes due to growth is insignificant compared with the rate of flux through the pools. This means that there is no accumulation of any of the metabolites in the cell i.e. $\mathbf{r} = 0$.

Equation 9 $0 = \mathbf{G}^T\mathbf{v}$

Some of the rates in vector **v** can be measured. The uptake rates of substrates and production rates of products and biomass components yield some of these rates. Matrix **G** and vector **v** can then be partitioned into a matrix and vector for the known rates and a matrix and vector for the unknown rates.

Equation 10 $\mathbf{0} = \mathbf{G}^T\mathbf{v} = \mathbf{G}_m{}^T\mathbf{v}_m + \mathbf{G}_c{}^T\mathbf{v}_c$

Where m = The matrix and vector whose rates can be measured
c = The matrix and vector whose rates must be calculated

If enough fluxes are measured the problem becomes determined or overdetermined and a solution can be found. If the problem is determined it can be solved for $\mathbf{v}_c$ by rearranging to give

Equation 11 $\mathbf{v}_c = -(\mathbf{G}_c{}^T)^{-1}\mathbf{G}_m{}^T\mathbf{v}_m$

If the system is overdetermined the G_c^T is not square and its inverse does not exist. In this case some of the measured fluxes can be treated as unknown, the system solved and the extra fluxes used to check the solution. Alternatively the pseudoinverse can be used. The pseudoinverse is given by

Equation 12 $(\mathbf{G}_c{}^T)^{\#} = (\mathbf{G}_c\mathbf{G}_c{}^T)^{-1}\mathbf{G}_c$

Equation 11 then becomes

Equation 13 $\mathbf{v}_c = -(\mathbf{G}_c{}^T)^{\#}\mathbf{G}_m\mathbf{v}_m$

The solution found is a least squares fit minimising discrepancy across the calculated rates. If all the measured fluxes are correct then there will be no discrepancies in the calculated rates. However as the measurement have associated errors and the measurement of biomass components is particularly prone to errors this is not an ideal solution. Data reconciliation methods have been devised which use the elemental composition of the biomass to help reduce errors in the macromolecular composition of the biomass (see Lange and Heijnen 2001). However as elemental data was not available these methods could not be used.

Whilst this method works well with data with little noise in the measured fluxes it does not work so well when there is a significant amount of noise in the data. This is because it assumes the measured fluxes are correct and then tries to fit the unmeasured fluxes to that data. The fitting should be taking place at the level of the measured fluxes making these consistent rather than the calculated fluxes. The difficulty of measuring biomass components more or less guaranteed a reasonable amount of noise in the data for this project. Tsai and Lee (1988) published a method which provides better estimates for both the measured and the calculated fluxes by reconciling the measured fluxes to each other so that they agree. In this method Equation 10 is reformulated as follows.

Equation 14

$$\begin{pmatrix} \mathbf{v}_m \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{G}_m^T & \mathbf{G}_c^T \end{pmatrix} \begin{pmatrix} \mathbf{v}_m \\ \mathbf{v}_c \end{pmatrix} = \mathbf{T}\mathbf{v}$$

Where

$\mathbf{I}$ = The identity matrix of equal dimension to the number of measured fluxes
$\mathbf{0}$ = The zero matrix
$\mathbf{G}_m$ = The measured fluxes from the G matrix partitioned into a new matrix
$\mathbf{G}_c$ = The calculated fluxes from the G matrix partitioned into a new matrix
$\mathbf{T}$ represents the new matrix formed by combining $\mathbf{I}$, $\mathbf{0}$ and $\mathbf{G}^T$.

On multiplication of $\mathbf{T}$ and $\mathbf{v}$ the top row yields $\mathbf{V}_m = \mathbf{V}_m$ and the bottom gives $\mathbf{0} = \mathbf{G}^T\mathbf{v}$. Matrix $\mathbf{T}$ is then partitioned into four smaller matrices.

Equation 15

$$\begin{pmatrix} \mathbf{v}_m \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} \\ \mathbf{T}_{21} & \mathbf{T}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{pmatrix}$$

This vertical partitioning is performed such that $\mathbf{T}_{21}$ and $\mathbf{T}_{22}$ contain the matrix $\mathbf{G}^T$. The horizontal partitioning is performed such that $\mathbf{T}_{22}$ is square and therefore can be inverted. $\mathbf{v}_2$ can now be found in terms of $\mathbf{T}_{21}$, $\mathbf{T}_{22}$ and $\mathbf{v}_1$.

Equation 16 $\mathbf{v}_2 = -\mathbf{T}_{22}^{-1}\mathbf{T}_{21}\mathbf{v}_1$

If this is substituted into Equation 15 and simplified then the following equation is obtained.

Equation 17 $\mathbf{v}_m = \mathbf{T}_r\mathbf{v}_1$

Where $\mathbf{T}_r = \mathbf{T}_{11} - \mathbf{T}_{12}\mathbf{T}_{22}^{-1}\mathbf{T}_{21}$

If the variance-covariance matrix $\mathbf{F}$ is known for the data then the maximum likelihood estimate of $\mathbf{v}_1$ can be calculated by the method of Madron *et al.* (1977).

Equation 18 $\hat{\mathbf{v}}_1 = (\mathbf{T}_r^T\mathbf{F}^{-1}\mathbf{T}_r)^{-1}\mathbf{T}_r^T\mathbf{F}^{-1}\mathbf{v}_m$

Where $\hat{\mathbf{v}}_1$ = The maximum likelihood estimate of $\mathbf{v}_1$

The maximum likelihood estimate for $\mathbf{v}_2$ ($\hat{\mathbf{v}}_2$) can now be calculated using Equation 16

Equation 19 $\hat{\mathbf{v}}_2 = -\mathbf{T}_{22}^{-1}\mathbf{T}_{21}\hat{\mathbf{v}}_1$

Now maximum likelihood estimates are available for both $\mathbf{v}_1$ and $\mathbf{v}_2$ which together make up $\mathbf{v}$ so both the calculated and the measured fluxes have been fitted to the measured data. This method gives much more realistic results than a simple least square solution.

The variance-covariance matrix used in the calculation above is derived from the variance of the measured data. This is not particularly easy to calculate for the measured fluxes. The individual data points have for most of the points been collected in duplicate or triplicate so variances can be calculated for the individual points. The data used in the model is not however these values but values arrived at by performing non-linear regression on the time series of these values and finding the slope of the line. This regression is a form of error minimisation designed to reduce the effect of the

variance of the individual points on the uptake rates. Calculating the variance on uptake rates derived in this way is not simple. For this project it was assumed that all the measured rates had proportionally the same variance. Variance was set at 1% for all uptake rates. This leads to the measured uptake rates being scaled to fit each other in a proportional manner which is important in a situation like this where the smallest rates are orders of magnitude smaller than the largest.

Covariance, the possibility of a mistake in measurement of one uptake rate causing a mistake in another was not considered as the software used did not support it.

2.8.2 Estimation of goodness of fit

When a variance weighted least squares method is used it generates not only the unknown fluxes but also re-estimates the known fluxes. It is then possible to measure how closely the estimated fluxes agree with the measured fluxes. The difference between the measured value and the estimated value is known as the residual. Residuals can be found for all the measured rates. These can then be combined to give a measure of the overall goodness of fit. The residuals are squared to make all positive, weighted according to the variance covariance matrix and then summed to give a value designated h. The larger the value of h the worse the fit.

The h value gives a quick estimate of the goodness of fit. It is not however a substitute for direct inspection of the flux distribution as its calculation is not based on a knowledge of the biochemical properties of the metabolic network and sometimes a reasonable h value can be calculated for an unreasonable flux distribution.

The values of h are χ^2 distributed. The h value can be compared with the χ^2 value for the number of degrees of freedom at a given confidence level (usually 0.95). Failure of this test suggests that there is a gross error in the data. The rogue rate can be identified by systematically leaving each of the measured rates as unmeasured, calculating the flux distribution and finding which rate has the biggest negative impact on the value of h.

χ^2 analysis was performed on all flux distributions calculated and if the test was failed gross error detection was performed to identify the incorrect rate and this rate was then removed.

2.8.3 Determination of condition number

The condition number is a measure of how well conditioned a matrix is. Software packages use numerical techniques to solve systems of linear equations. This involves many calculations which will involve rounding errors. A poorly conditioned matrix may amplify these errors leading to a grossly inaccurate result. The condition number is calculated as follows

Equation 20 $$C(G^T) = \left\| G^T \right\| \left\| (G^T)^{\#} \right\|$$

2.9 Macromolecular Composition

Limited data was available for the compositions of macromolecules in *Saccharopolyspora erythraea.* Compositions were constructed from the best sources available making use of bioinformatics databases. The DNA of *S. erythraea* is 76.9% GC rich (Labeda 1987). From this the DNA and mRNA compositions were found. The rRNA composition was estimated using sequences available on the rRNA www server (http://rrna.uia.ac.be/index.html). Small subunit information was from *S. erythraea.* Large subunit information was not available for *S. erythraea. Mycobacterium tuberculosis* small subunit rRNA was found to be closer to *S. erythraea* than that of any other species for which large subunit information was available. For this reason *M. tuberculosis* large subunit information was used in the model. The tRNA composition was based on sequences available for *M. tuberculosis* from the Genomic tRNA Database (http://rna.wustl.edu/tRNAdb/).

Protein composition was estimated using data from the *Streptomyces coelicolor* genome project (Bentley *et al* 2002). Amino acid compositions of all the open reading frames in the database were averaged to give an average amino acid composition for protein in the cell. This does not take into account the relative abundance of the different proteins which is not known. The analysis relies on the levels of protein expression not being linked to the amino acid composition. A best guess of fatty acid composition was made based on Goodfellow *et al* (1989), Wallace *et al.* (1995) and Gamian *et al.* (1996). A best guess of phospholipid composition was made based on LeChevalier *et al* (1977). Proportions of phospholipid and triacylglycerol in the total lipid was estimated based on

Olukoshi and Packter (1994) and on the proportions of lipid synthesised before and after nitrate depletion. Peptidoglycan composition was based on Goodfellow *et al.* (1989). The amount of peptidoglycan was not measured but was assumed to make up 5% of the DCW based on data reported by Davidson (1992). Arabinogalactan composition was measured after hydrolysis and was found to be in reasonable agreement with that found for *M. tuberculosis* (Crick *et al.* 2001). Teichoic acids are not found in members of the suborder Pseudonocardineae (Naumova *et al* 2001) and are therefore not included. Mycolic acids are not found in *S. erythraea* (Labeda 1987) and are therefore not included in the model. Relevant information on metabolic pool sizes was not available from the literature for this organism and this fraction of the biomass was neglected in modelling. During the course of this work *S. erythraea* was found to produce a capsule in the stationary phase (see section 3.3.4). The composition of the capsule was unknown and it was not possible to quantify the amount so this was neglected in the modelling.

2.10 Metabolic Network Construction

The metabolic network used here was based on that of Daae and Ison (1999). To this a large number of pathways were added. A full list of the reactions included in the network and the abbreviations used in the network can be found in Appendix 1. A schematic diagram of the metabolic network used is presented as Figure 7. This diagram is not exhaustive. All reactions of small molecule metabolism used in the model are included. Macromolecule synthesis is not included. The interconnectivity of the network is simplified to make the diagram more readable.

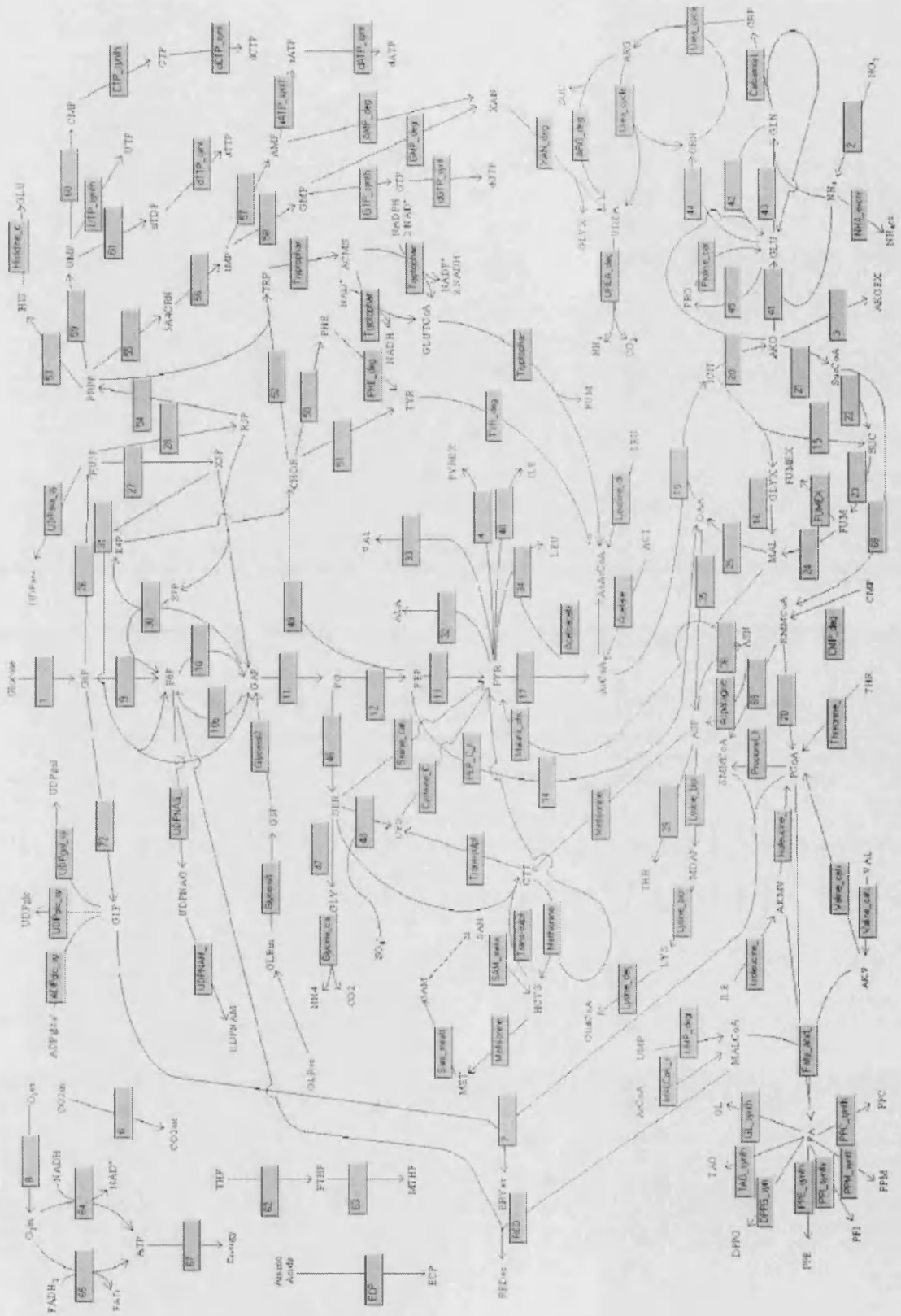


Figure 7: Diagrammatic representation of the metabolic network used in this work. For full reactions and abbreviations see Appendix 1.

2.10.1 Biomass synthesis

The model proposed by Daae and Ison (1999) dealt with growth of biomass as one large reaction going from precursors of macromolecules to biomass. This meant that the proportions of the biomass components could not be changed during the course of the fermentation. Experiments performed in this project clearly showed that the proportions of the biomass components do change especially during the stationary phase. To improve the model it was necessary to allow the synthesis of the macromolecular components of the biomass to be varied independently of the dry cell weight readings. To this end the biomass reaction was replaced with individual reactions for the synthesis of each of the macromolecules. The Daae model used biomass information derived from *E. coli.* For this model the choice of biomass components and the monomeric makeup of these components was based on data from *S. erythraea* or related organisms (see section 2.9). Energy requirements for polymerisation of DNA, RNA and protein were based on the values given in Stephanopoulos *et al* (1998).

2.10.2 Amino Acid Catabolism

Pathways for catabolism of all the natural amino acids were added. Valine, leucine and isoleucine are assumed to be catabolised by the standard branched-chain amino acid pathways (Voet and Voet, 1990), this is the case in the closely related *streptomyces* (Hodgson, 2000). The first step in the pathway produces branched chain α-keto acids. The α-keto acids of valine and isoleucine are precursors for iso and anteiso-fatty acid synthesis (Oku *et al.* 1998) and can also be broken down further to produce propionyl-CoA a precursor of erythromycin. Leucine catabolism leads to acetoacetate which feeds back into central metabolism via acetate. Threonine catabolism is assumed to be via threonine dehydratase leading eventually to propionyl-CoA, as is seen in the *streptomyces* (Hodgson, 2000). Little is known about serine, glycine and cysteine catabolism in *saccharopolyspora* or related genera. Their catabolism is assumed to follow the standard pathways (Voet and Voet, 2000). The enzymes of alanine formation and aspartic acid formation are reversible therefore no new pathways were required for their degradation. Phenylalanine and tyrosine catabolism where seen in the

streptomyces proceeds via the homogentisate route to acetoacetate and fumarate (Hodgson, 2000) and the pathway as described in Voet and Voet (1990) is used here. Tryptophan catabolism in the *streptomyces* occurs via hydroxy-anthranilate (Hodgson, 2000) to acetoacetate and alanine (Voet and Voet, 1990). Histidine is assumed to be catabolised by the standard pathway (Voet and Voet, 1990) as is the case in the closely related *streptomyces* (Hodgson, 2000). Lysine catabolism is thought to proceed via cadaverine yielding acetoacetate in the *streptomyces* (Fothergill and Guest, 1977; Hodgson, 2000) and is assumed to do so in *saccharopolyspora* too. Proline and asparagine catabolism are assumed to be by standard pathways (Voet and Voet, 1990) as is the case in the *streptomyces* (Hodgson, 2000). Arginine in the *streptomyces* is catabolised to succinate (Padilla *et al.* 1991; Hodgson, 2000), it is assumed to be the same in *saccharopolyspora.* Methionine catabolism is assumed to proceed by the regular pathway (Voet and Voet, 1990).

2.10.3 Nucleotide Metabolism

Pathways for catabolism of nucleotides were added. CMP and UMP are assumed to be catabolised to malonyl-CoA (Voet and Voet, 1990) the little evidence available in the *streptomyces* supports this (Hodgson, 2000). TMP is not considered because DNA catabolism is not usually seen in *S. erythraea* (see results sections). AMP catabolism is assumed to proceed via inosine to xanthine where it converges with GMP catabolism leading to urea and glyoxylate formation (Voet and Voet, 1990) as is the case in the *streptomyces* (Hodgson, 2000).

2.10.4 Cofactors

s-Adenosylmethionine methylation plays an important role in erythromycin production and the standard pathway for its formation and regeneration has been included (Voet and Voet, 1990). NADPH also plays an important role in the production of erythromycin. A number of enzymes produce NADPH and this has been studied in some detail in *S. erythraea* by Roszkoski *et al.* (1971). They found NADPH production at three sites not included by Daae (1999). These were isocitrate dehydrogenase glutamate dehydrogenase and malic enzyme. Malic enzyme has been included in the

model, however it is assumed to be inactive under the conditions employed because it is incompatible with phosphoenolpyruvate carboxylase. NADPH dependence was introduced for both isocitrate dehydrogenase and glutamate dehydrogenase. In the stationary phase however isocitrate dehydrogenase was modelled as NADH dependant in order to prevent the model calculating thermodynamically unfavourable flux distributions. This is discussed in more detail in section 4.3.4.1.2.

2.10.5 Anaplerotic Pathways

A number of pathways are capable of providing for the anaplerotic requirements of the cell. Phosphoenolpyruvate carboxylase was inherited from the model of Daae and Ison (1999). This is thought to be the main anaplerotic pathway in steptomycetes (Dekleva and Strohl 1988a, Hodgson 2000). Although included in the model the glyoxylate shunt has not been used in these studies as lipids have not been used as carbon source. An alternative anaplerotic route suggested by Han and Reynolds (1997) can be ruled out as crotonyl-CoA reductase is not present in *S. erythraea* (Liu and Reynolds 2001). Likewise pyruvate carboxylation is ruled out by the absence of pyruvate carboxylase activity (Roszkowski *et al* 1971).

2.10.6 Gluconeogenesis

Phosphoenolpyruvate carboxykinase was added to the model because in the late stationary phase after glucose runs out α-ketoglutarate is taken up and has to be converted to pyruvate to enter the TCA cycle. Fructose-1,6-bisphosphatase has been added because during the exponential phase fluxes through the pentose phosphate pathway are so high that a cycle is formed driving the early steps of glycolysis backwards. Driving the reaction catalysed by phosphofructokinase backwards is thermodynamically unfavourable and fructose-1,6-bisphosphatase performs this role instead (Voet & Voet 1990). When modelling only one of these reactions is included in order to prevent linear dependencies in the stoichiometric matrix.

2.10.7 Uptake and secretion

The work of Ushio (2003) showed that *S. erythraea* can utilise glycerol. Based on this the pathways of glycerol uptake were included. Glycerol was not however eventually utilised as a carbon source in this work.

Low levels of ammonia secretion have been seen during the transition from exponential to stationary phase. A flux for this has been included and is assumed to occur by diffusion through the cell envelope. The transition phase however proved too dynamic to model with flux balance analysis. This pathway was however used in the stationary phase because the nitrogen balance did not close. Nitrogen produced by the breakdown of protein had no known sink. This pathway was provided as a theoretical sink for this nitrogen. Without this sink the model produced unsatisfactory flux distributions as it can only balance the network by ignoring the data on protein and RNA degradation.

Protein secretion has been measured and included in the model. Whilst the energetic cost for formation of extracellular protein is the same as for cellular protein there is also a cost associated with translocating it across the cell membrane. The mechanisms involved are not fully elucidated and exact energy costs are not known. Van Wely *et al* (2001) reported that SecA is responsible for translocation of 90% of excreted proteins in *B. subtilis* and the same system is also present in *mycobacteria* and *streptomyces*. SecA is an ATPase and Karamanou *et al* (1999) reported that each hydrolysis of ATP drives 20-30 amino acids through the membrane. This however may not be the only energetic cost as they report that the proton motive force also seems to be involved. The extra energetic cost of translocating the extracellular protein has been included in the model.

During the course of the project the secretion of fumarate in the stationary phase was discovered (see section 0). A pathway for its secretion has been included in the model. It is assumed that fumarase is passively transported across the cell boundary.

The pathway for erythromycin A production has been included. This can be considered in two parts, synthesis of the polyketide moiety and synthesis of the two attached sugar moieties. The synthesis of the polyketide is well understood and the pathway has been reviewed by Katz (1997). The pathway used here has been shown in detail in section 1.2.1.2. The reactions connecting succinate in central metabolism to propionyl-CoA

and Methylmalonyl-CoA which provide the precursors for the synthesis of the polyketide were taken from a model proposed by Ushio *et al* (2003). The synthesis of the sugar moieties has only recently been elucidated and there are still minor disputes over the pathway. The pathway of Salah-Bey *et al* (1998) has been used here. This pathway has been shown in detail in section 1.2.2.2.

The pathway for red pigment formation has been included. The details of the biosynthesis of this compound are still being worked out but the general structure is provided by Cortés *et al.* 2002. The reduced cofactor requirement was estimated using redox analysis of the products, intermediates and precursors and was presumed to be supplied by NADPH

2.10.8 Storage Compounds

Trehalose synthesis in *streptomyces* follows the pathway elucidated by Elbein, reported in Hodgson (2000), it is assumed to follow the same pathway in *S. erythraea.* Trehalose catabolism is assumed to proceed via hydrolysis to glucose as is the case in the *streptomyces* (Aisaka & Masuda, 1995). Glycogen formation was assumed to follow the standard route utilising ADP-glucose as is the case for *S. coelicolor* (Martin *et al* 1997). It was later however removed as no firm evidence was obtained that glycogen made up a significant fraction of the biomass see section 2.6.11. Lipid synthesis pathways were based on the lipid compositions from Goodfellow *et al* (1989) LeChevalier *et al* (1977) and Wallace *et al.* (1995). The pathways of lipid biosynthesis are based on Stryer (1995), Ratledge & Wilkinson (1988), Wallace *et al.* (1995) and Oku *et al.* (1998).

2.10.9 Cell Wall Components

The synthesis of UDP-N-acetylglucosamine and UDP-N-acetylmuramate is assumed to be the same as in *E. coli.* Reactions were take from the model of Edwards and Palsson (2000). The formation of peptidoglycan is assumed to follow the usual pathway (meso-diamino-pimelate branch) (KEGG database http://bioinfo.weizmann.ac.il:3456/kegg) with the exception that glycine is substituted for L-alanine in line with the type IV cell wall composition seen in this genus (Goodfellow et al 1989). Synthesis of UDP-

arabinose is assumed to proceed via the pentose phosphate pathway as in the *Mycobacteria* (Scherman *et al.* 1995). Synthesis of UDP-galactose is assumed to proceed by the standard pathways (see KEGG database). The pathway for polymerisation is based on experimental results regarding arabinose to galactose ratios and on data reported by Crick *et al* (2001). Branching of the arabinose chains has been neglected.

2.10.10 Selection of Pathways

The pathways in this model were not all used at once. Some of the pathways are mutually exclusive because if both are used at once linear dependencies arise in the stoichiometric matrix. This occurs because the modelling method is not able to distinguish between the two pathways. This means that although it can calculate the combined flux through both pathways it can't distinguish how much is attributable to each pathway. Examples of such reactions include the reactions of anaplerosis. If more than one anaplerotic pathway is used at once linear dependencies occur. Likewise futile cycles such as that involving phosphofructokinase and hexose-diphosphatase cause linear dependencies. Unfortunately some linearly dependent pathways are large and complex involving much of metabolism. For example if a transhydrogenase is introduced to the metabolic network used in this work the TCA cycle and the pentose phosphate pathway become indistinguishable and it becomes impossible to calculate flux distributions through a large proportion of central metabolism. Other pathways were selected in order to restrict the model according to the biochemistry of the pathways. For example to the model there is no inherent difference between the pathways used to synthesise amino acids and those used to degrade them. In order to ensure the use of the correct pathway the pathways available were restricted according to whether protein was being produced or broken down in the cell. A little care is needed with this as some amino acids are used in lipid synthesis and cell wall synthesis and so may be required in the stationary phase in excess of the amounts produced by protein degradation.

2.10.11 Condition number

Numerical techniques are often used to solve matrix problems. These techniques can cause large rounding errors if the matrix is ill-conditioned (see Stephanopoulos *et al.* 1998 page 348). To identify whether this is likely to be a problem the condition number of the matrix can be calculated. The larger the condition number the more accurate data measurements have to be to achieve a correct answer. Given that there is a limit on how accurate biochemical measurements can be there is a limitation on the condition number for accurate calculation of stoichiometric matrices. Stephanopoulos *et al.* (1998) hold that measurements should be carried out with a precision that carries the same number of digits as the condition number. Whilst some analysts may claim to be able to reliably perform biochemical assays to ± 1% many important biochemical assays cannot come close to this especially biomass composition estimations (see notes on assaying protein levels section 2.6.4). This would restrict the condition number to being less than 100 this causes problems for some larger models. Savinell and Palsson (1992) consider that a condition number of over 1000 is undesirable for flux balance analysis. The condition number of the matrix used in this work is 179 which is greater than that advised by Stephanopoulos *et al.* (1998) but well within the limit set by Savinell and Palsson (1992).

3 BIOMASS COMPOSITION AND ITS IMPACT ON FLUX BALANCE ANALYSIS

3.1 Introduction

One of the most fascinating aspects of actinomycetes such as *S. erythraea* is the developmental processes they undergo. On solid media this is seen in the formation of aerial mycelia sporulation and production of secondary metabolites. Under the conditions used for submerged fermentation aerial mycelia and sporulation are not seen however changes in the macromolecular composition of the biomass do occur. For example Packter and Olukoshi (1995) found that during the stationary phase *S. coelicolor* could accumulate lipid up to 80% of the cell volume.

Flux balance analysis was chosen as the method to be used in this project to model the primary and secondary metabolism of *S. erythraea*. This method determines the flux of metabolites through all the pathways in a metabolic network based on measured inputs and outputs. It is however necessary to consider whether it is applicable to an organism that displays large changes in biomass composition. If it is applicable principles for its application in this situation need to be considered.

Flux balance analysis works by calculating a flux distribution for the cell based on the uptake of nutrients, secretion of products and synthesis of biomass components by the cell. The usual way flux balance analysis is applied is to assume that the biomass composition of an organism remains constant. The rate of synthesis of the biomass components is then based on the growth rate of the organism. Since it is difficult to accurately measure some biomass components the composition is often based on that of model organisms such as *E. coli* where this data is known with some accuracy. These assumptions facilitate application of flux balance analysis to a wide range of organisms as fewer and simpler measurements are needed. However given that the biomass compositions of organisms such as *S. erythraea* can show large changes questions remain over the validity of using a constant biomass composition for such organisms.

Daae and Ison (1999) studied the primary metabolism of *Streptomyces lividans* during exponential growth using flux balance analysis. They performed a sensitivity analysis on their calculations to reveal the impact on the flux distribution of a 20% change in demand for each of the monomeric components required to synthesise biomass. The analysis revealed that the changes to flux distribution in the network were generally insignificant. This however does not justify ignoring biomass composition in the case of organisms performing secondary metabolism. As Daae and Ison clearly realised it is not so much the biomass composition as the flux of monomeric components to biomass that is important. Daae and Ison's study was undertaken in an organism that performs secondary metabolism but was a study of its primary metabolism during the growth phase. During the growth phase the biomass composition is reasonably constant and so the flux distribution of precursors to biomass is reasonably constant. Their interest was in what difference it would have made if a slightly different constant biomass composition was being produced in the growth phase. This allowed the justification of using the biomass composition of another species to calculate flux distributions in this species. This kind of change in biomass composition is very different to the kind of change occurring when the growth phase turns into the stationary phase and secondary metabolism occurs. The move from growth phase to stationary phase is usually triggered by nutrient limitation. Some of the change in biomass composition is a direct consequence of the nutrient limitation. If the nitrogen source becomes depleted the synthesis of protein and RNA halts immediately due to the stringent response (Chatterji and Ojha 2001). Protein and RNA levels generally then begin to decline. This means that in terms of some of the monomeric fluxes to biomass there is a change of at least 100% even though the measured biomass composition may only be changing slowly as some components continue to accumulate and others decline.

In order to investigate these changes further a study of the biomass composition of *S. erythraea* during batch growth on nitrate limited media was undertaken. To the best of my knowledge this is the first such study of *S. erythraea*. Using this data the impact of assuming a constant biomass composition on flux balance analysis was investigated. This section also contains a brief description of the process of developing the analytical techniques used to quantify the biomass composition. There is also a preliminary description of the presence of a capsule not previously reported in *S. erythraea*.

The main thrust of this chapter will be an attempt to show that it is not valid to assume that biomass composition is constant when performing flux balance analysis of *S. erythraea* in the stationary phase. If this is so there will be a significant difference in flux distributions achieved using a constant biomass composition rather than a dynamic biomass composition. Moreover the flux distribution calculated using the dynamic biomass composition will be more consistent with the measured data than the flux distribution calculated assuming a constant biomass composition.

3.2 Materials and Methods

All materials and methods are as described in chapter 2 except where otherwise indicated during the section on development of techniques. *S. erythraea* white variant wild type NRRL 2338 was used throughout this section. Fermentation was performed in 20 litre LH fermenters as described previously (see section 2.3). Methods used for investigating the formation of capsular material are described in section 3.3.4.2.

3.2.1 Calculations

Flux balance analysis was performed as described in section 2.8. Two different methods of describing the biomass composition were used. One used an average composition of the biomass derived from the exponential phase when biomass composition is constant. This is referred to as the static model. The other method used rates of formation of the individual biomass components throughout the fermentation. This is referred to as the dynamic model.

In order to compare the results calculated in this section with those found by Daae and Ison (1999) the results were assessed in the same way as done in their work. Flux distributions were calculated using both the dynamic and the static model. The absolute difference between each flux calculated with the two methods was found. This was then calculated as a percentage of the absolute value of the flux as found by the dynamic method. This was done for the entire flux distribution. The mean of these values was found both for the whole network and for the subset defined as central metabolism. The

maximum of these values was also found for the whole network and for central metabolism.

3.3 Results and Discussion

Before flux balance analysis can be meaningfully applied to comparison of fermentations consistent behaviour of the biomass must be achieved. Early fermentations were inconsistent and often displayed unusual behaviour. Eventually two main factors causing this were identified. The first was the quality of the inoculum. It is important that the inoculum is rapidly growing at the time of inoculation. It is also important that the amount of biomass used to inoculate is not too small. Too low an inoculum leads to lower growth rates. A 10% inoculum with a dry cell weight of 3-5 g/l was found to give the best results. The other factor was the choice of antifoam. Polypropylene glycol was used initially but was found to reduce the growth rate. It also interfered with some of the assays as at higher concentrations it can form an emulsion under certain conditions which is hard to remove. This emulsion interferes with spectrophotometric assays and makes microscopic inspection of cultures difficult as the particles appear of similar size and shape to cocci. M. Ushio (2003) investigated the effect of antifoams further and concluded that silicon based antifoams were best for used with *S. erythraea.* For this reason Dow Corning "DB-110 A" a silicon based antifoam was used in later experiments. The data from the earlier experiments has not been used for flux analysis here due to unreliable growth. However methods for measuring biomass composition could still be tested and initial ideas of the biomass composition formed.

In order to test whether measuring the changing biomass composition allowed more accurate calculation of flux distributions it was necessary to measure the biomass composition of *S. erythraea* many times throughout the fermentation. Simple reliable methods of measuring these components were therefore required. As the biomass components of *S. erythraea* had not been studied in detail the biomass composition of *E. coli* was used as a guide of what to look for. In *E. coli* the major biomass constituents are protein (52%), carbohydrate (17%), RNA (16%), lipid (9%) and DNA

(3%) (Stouthamer 1979 quoted in Stephanopoulos *et al.* 1998). It was assumed that most of the carbohydrate would be in the form of glycogen.

3.3.1 Development of Analytical Techniques

Techniques for determining cellular DNA, RNA, lipid and glycogen were developed first. The methods used were as stated in section 2. The orcinol method was used for RNA determination. The results of one of the early fermentations are presented as Figure 8. The DNA and RNA profiles exhibit reasonable shapes that seem to follow the dry cell weight. The lipid profile contains too much noise to be useful. The glycogen assay did not produce useful data and the profile is not included. Two profiles are presented for the dry cell weight. One is the raw data and the other has been adjusted using the DNA data (see below).

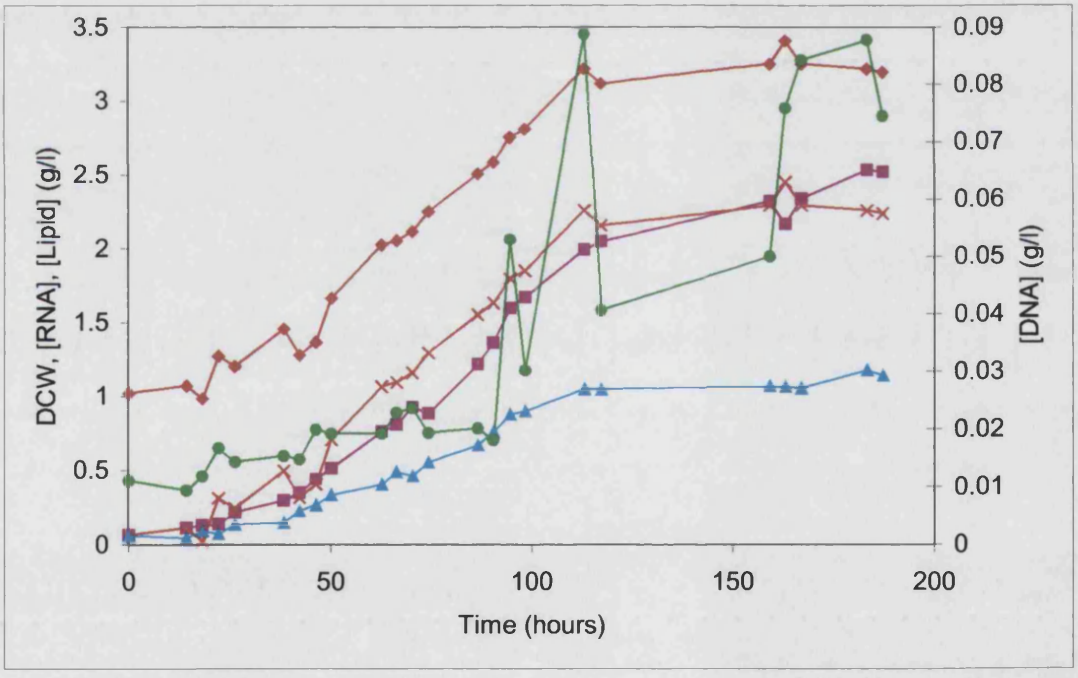


Figure 8: White variant wild type biomass composition profile fermentation 1. Dry cell weight ◆, Adjusted dry cell weight ×, DNA ▲, RNA ■, Lipid ●. Adjusted dry cell weight was calcualed based on Figure 9 below.

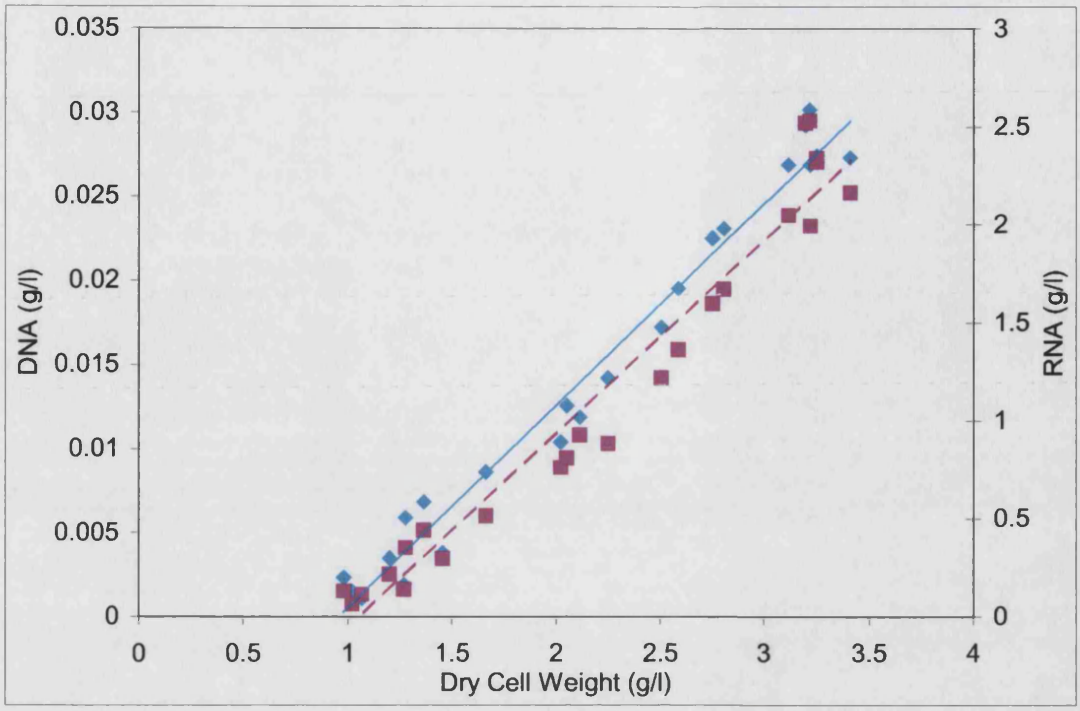


Figure 9: Dry cell weight and nucleic acid comparison. DNA ◆ , RNA ■ . This figure shows that there is a systematic error of about + 1 g/l on the dry cell weight measurements.

3.3.1.1 Dry Cell Weight

The DCW profile in Figure 8 is not as expected. The growth phase appears linear and the initial dry cell weight measurement was too high based on the DCW of the inoculum (1g/l). (The low DCW in the inoculum is thought to be one of the causes of the rather slow and unusual growth seen in this fermentation.) When the natural logarithm of the DCW was plotted against time it confirmed that the growth phase was not exponential.

These discrepancies pointed to a problem with the DCW measurements. Failure to completely remove residues of the media from the filter would lead to a systematic error in the DCW readings. To check for this DNA and RNA content were plotted against DCW. Assuming that in the exponential phase they form a constant proportion of the biomass a linear relationship would result. With no systematic error the line should pass through the origin. If there is a systematic error the size can be found as the intercept on the x axis. The results shown in Figure 9 clearly show that there is such a systematic error. Both the DNA and the RNA content point to an error of around 1g/l. The value derived from the DNA content was used to calculate the true DCW by

subtraction of the extra mass. The adjusted DCW readings did follow the expected exponential pattern in the growth phase. This method of checking and adjusting the DCW was used on all future fermentations. The problem was smaller when filters were dried for 24 hours in an oven rather than using the moisture analyser (see section 2.6.1 for the drying methods). This suggests that the moisture analyser may not be entirely drying the filters.

3.3.1.2 DNA

The DNA method used gave a reasonable profile both in terms of shape, it followed the adjusted DCW and in terms of amount. DNA made up around 1% of the DCW of the cell at all stages in the fermentation. The only difficulty found with this method was that the standard curve is not linear. The DNA concentration was calculated from a cubic function fitted on a plot of OD_{600} Vs concentration of DNA. Reversing the axes in this way makes calculation of DNA concentrations simpler.

3.3.1.3 RNA

The RNA profile conforms to expectations. It looks quite like the DNA profile in shape however the levels of RNA found imply that RNA makes up 80-100% of the DCW of the cell. This is extremely unlikely. No problem could be found with either the method or the calculations. The method was repeated on a number of fermentations and the results were consistently within this range. The orcinol method detects pentose sugars hence its use for RNA determination. It was concluded that there must be significant amounts of another pentose sugar in the cell. Taxonomic references such as Labeda (1987) referred to galactose and arabinose as being present in the cell wall but with no quantification of the levels. Arabinose is a pentose sugar. A method for quantifying the galactose and the arabinose was devised see section 2.6.6. Arabinose in the form of arabinogalactan, a cell wall polymer, was found to make up more than 10% of the DCW hence its strong interference with the orcinol method. The orcinol method was therefore abandoned and a method based on the UV absorption of RNA (Benthin, 1991) was used instead.

The Benthin method is, despite the paper's claims to the contrary, more complicated and time consuming than the orcinol method. The results it yields for *S. erythraea* are however good and in line with the results found in other related genera and it was adopted in later experiments (see section 3.3.3.1 for RNA results using the Benthin method).

3.3.1.4 Lipid

The lipid profile in Figure 8 above shows a high degree of scatter. However with practice and refinement the method was found to give a reasonable profile. The main problems with the method are two fold. Firstly standardisation, this is a colorimetric assay and different lipids give different strength signals. Without a representative standard the results can only be treated as rough estimates. Secondly the method uses the pellet from a complicated multistep extraction process. Each step of the process provides opportunities for biomass to be lost from the pellet. To counter this the extraction process has to be performed with great care.

Zöllner and Kirsch (1962) noted that the vanillin assay gives different responces for different types of lipid. However given that no standard for *S. erythraea* lipid was available triolein was used. Triolein exhibits a fairly strong response in this assay so the results probably underestimate the lipid content of the biomass slightly.

3.3.1.5 Glycogen

The glycogen assay was performed in all early experiments. The results were inconsistent and usually only very low concentrations were found. The method is not very specific. It relies on extraction and precipitation steps to remove the glycogen from other sugars in the cell. As the work proceeded it became clear that there were other carbohydrates in the cell including arabinogalactan, a polysaccharide and trehalose a disaccharide. These sugars made up a large fraction of the dry cell weight of *S. erythraea*, more than the carbohydrate fraction of *E. coli*. One profile which did seem to show low levels of glycogen was compared with the arabinogalactan profile for the same fermentation and it was found to have the same shape. HPLC analysis of the

hydrolysed extracts did not show any glucose. From this it was concluded that glycogen did not form a significant proportion of the biomass in *S. erythraea*. The glycogen assay was not performed in later experiments. Given time an enzyme assay based on amylase might be able to confirm the absence of glycogen however this line of investigation was not pursued.

3.3.1.6 Intracellular Protein

Protein was measured by the Bradford assay. However in order to use this method it is necessary to make the protein available to the reagent. Two methods were tried for this: sonication and chemical extraction, the results are presented in Figure 10. The two protein profiles seem to be roughly in agreement in the early growth phase and late exponential phase. However during the transition phase the protein levels measured by the sonication method increased to considerably more than the levels found by alkaline extraction. The reason for this is unknown but it was repeatable. The shape of the profile for sonicated cells was closer to that expected. The peak protein level found by this method was however considerably higher than the peak dry cell weight. The alkaline method gave protein content at about 50% of the dry cell weight. This was much closer to the value found by Davidson (1992) for *S. coelicolor*. For this reason the alkaline extraction method was used in all later experiments. The Bradford assay is pH sensitive. For this reason standards should be made in the same alkaline solution as used for the extraction. It is best however to dissolve the BSA in water first and then adjust the pH as powered BSA denatures in alkaline solution to form a gel which is hard to dissolve.

The Bradford assay works by the binding of the Coomassie blue dye to protein molecules. However the dye binds particularly strongly to basic amino acids. This means that the amino acid composition of the protein influences the response of the assay, it also seems to be effected by the size of the protein (Sapan *et al.* 1999). This makes the assay appropriate only when representative standards are used. However there is no representative standard for *S. erythraea* protein. This is not only a problem for the Bradford assay, other simple colorimetric assays fair no better. Bio-Rad Laboratories GmbH publish a table of protein's responses with their Bradford assay

reagent from which it is possible to calculate the strength of response of various common proteins compared to an "average" protein. Using this information bovine serum albumin was calculated to produce a response 2.1 times as strongly as average protein. Bovine serum albumin was then used as the standard for *S. erythraea* protein. *S. erythraea* protein was assumed to produce an average responce and the concentration of protein calculated based on this.

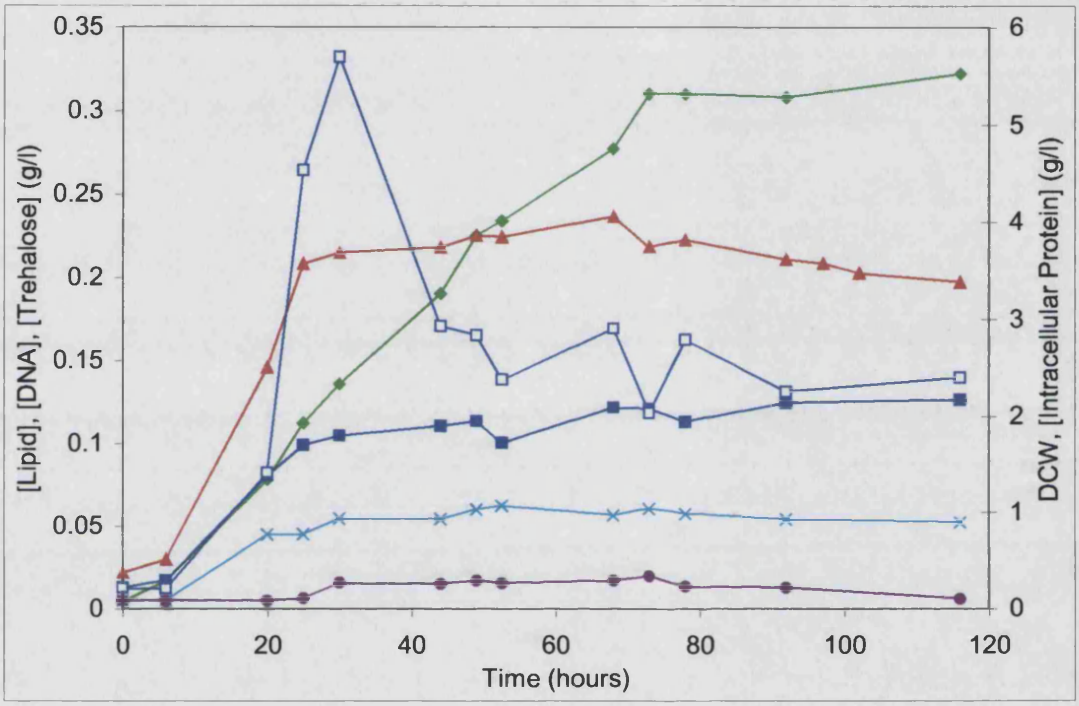


Figure 10: White variant wild type biomass composition profile fermentation 2. Lipid ◆, DNA ×, Dry cell weight ▲, Intracellular protein by alkaline extraction ■, Intracellular protein by sonication , Trehalose ●.

3.3.1.7 Trehalose

A second fermentation also provided early profiles for trehalose and arabinogalactan. The biomass composition from this fermentation is presented in (Figure 10). Trehalose accumulation has been noted in *streptomyces* species (eg Braña *et al.* 1986). In order to investigate its presence in *S. erythraea* a suitable method had to be developed. An enzyme assay can give a very high degree of certainty regarding the presence of compounds like this. They are however quite difficult to devise from scratch. In this case the enzyme assay would produce glucose for analysis which, given that the carbon source is glucose, is not ideal as great care would have to be taken to prevent

interference. It is also possible to detect trehalose directly by HPLC. A method for detecting sugars was already in use. The method is designed to resolve mixtures of mono and disaccharides. With appropriate extraction the trehalose could be detected using this method with only a change in the injection volume. The trehalose peak was identified by comparison with a standard. When spiked with the standard the peaks were irresolvable. A first attempt at quantifying the trehalose in biomass seemed to show the presence of very low levels in the biomass (see Figure 10). However in later experiments showing more usual growth patterns higher levels were seen especially in the stationary phase (see Figure 13).

3.3.1.8 Arabinogalactan

Most taxonomic papers dealing with the genus *Saccharopolyspora* mention the presence of arabinose and galactose in the cell wall (eg Labeda 1987). They do not however report the quantity or form taken by the sugars. The closely related *M. tuberculosis* contains the sugars in the form of arabinogalactan, a polysaccharide in its cell wall on to which mycolic acids are attached (Crick *et al.* 2001). It became apparent during measurements of RNA concentration that significant quantities of a pentose sugar other than ribose or deoxyribose were contained in the biomass (see section 3.3.1.3). The orcinol method for RNA measurement uses a two step extraction procedure to extract the RNA. The interfering sugar was found in the same fraction as the RNA. This showed that it was not in a soluble form in the cell. All soluble compounds are removed in the first extraction step. It was however hydrolysed by the mild acidic conditions used in the second step. This pointed to arabinogalactan as the source of these sugars. Arabinose and galactose can be measured using a standard HPLC method for sugars. The extract was however too acidic for the column and was therefore neutralised with NaOH prior to injection. The arabinose and galactose were identified by comparison with standards. When spiked with the appropriate standard the peaks co-eluted with their respective standards and could not be resolved.

3.3.2 Monomeric Composition

In order to have an accurate picture of the flux of biomass precursors to biomass it is not enough to just know how much of each of the macromolecules is being produced. It is

also necessary to know the proportions of the precursors used for the synthesis of these macromolecules. The monomeric composition of some of the macromolecules were determined for this work, others depended upon literature data (see section 2.9).

3.3.2.1 DNA

The monomeric composition of DNA was calculated based on the GC content of 76.9% reported by Labeda (1987). The monomeric composition is given in Table 4.

Table 4: Monomeric composition of DNA

Base	Frequency (%)
Adenine	11.55
Guanine	38.45
Cytosine	38.45
Thymine	11.55

Based on this the average relative molecular mass (RMM) per base for *S. erythraea* DNA was calculated as 307.9 Daltons per base.

3.3.2.2 RNA

RNA is found in three main forms in the cell. Messenger RNA (mRNA), transfer RNA (tRNA) and ribosomal RNA (rRNA).

The monomeric composition of mRNA is different for each gene transcribed. It is however dependant on the composition of the DNA it is transcribed from and can be calculated from this. This assumes that the coding DNA has the same GC bias as the non-coding sections. The composition calculated on the basis of this is presented in Table 5.

Table 5: mRNA monomeric composition

Base	Frequency (%)
Adenine	11.55
Guanine	38.45
Cytosine	38.45
Uracil	11.55

The monomeric composition of rRNAs and tRNAs cannot be expected to follow the composition of the DNA. They can however be derived directly from published tRNA and rRNA sequences. Whilst there are a large number of different tRNAs in any organism they are all reasonably similar. Those published have been taken as representative of those not published. There are three rRNA molecules per ribosome: 23 S, 16 S and 5 S. As the 5 S rRNA represents only 2% of the RNA in the cell its impact on the monomeric composition of the RNA has been neglected.

There is only limited sequence information available for *S. erythraea*. Information on 16S rRNA is however available for a very large number of organisms due to its use in taxonomic studies. The 16 S rRNA composition of *S. erythraea* was compared with those from *Streptomyces*, *Rhodococcus* and *Mycobacterium* species. Of these the *Mycobacteria* showed the closest composition to that of *S. erythraea*. Where RNA sequence data was not available for *S. erythraea*, *Mycobacteria* data was used instead. The details of the method used are given in section 2.9.

Information for the 16 S rRNA was found for *S. erythraea* and is presented in Table 6.

Table 6: 16 S rRNA monomeric composition

Base	Frequency (%)
Adenine	21.2
Guanine	23.4
Cytosine	35.5
Uracil	19.9
Number of bases	1274

Information for the 23 S rRNA was based on the average of 4 sequences from *M. tuberculosis*. The resulting monomeric composition is given in Table 7.

Table 7: 23 S rRNA monomeric composition

Base	Frequency (%)
Adenine	24.2
Guanine	23.7
Cytosine	32.7
Uracil	19.4
Number of bases	3125.5

The monomeric composition of the tRNA was based on the average of 6 sequences from *M. tuberculosis*. The resulting composition is presented in Table 8.

Table 8: tRNA monomeric composition

Base	Frequency (%)
Adenine	17.1
Guanine	30.5
Cytosine	35.5
Uracil	16.9

Neidhart *et al.* (1987) determined the proportions of the different RNA types in the RNA pool of *E. coli* (see Table 9). These proportions have been used here as proportions for more closely related organisms were not found.

Table 9: *E. coli* RNA pool

Type of RNA	Fraction of RNA pool
rRNA	81.5%
tRNA	14.5%
mRNA	4.0%

From the tables above an average RNA monomeric composition was calculated by multiplying the frequency of each base in a type of RNA with the fraction of that type of RNA in the pool and then summing the results. The frequencies of the bases in total RNA is presented in Table 10. The average relative molecular mass (RMM) per base of *S. erythraea* RNA was calculated as 323.1 Daltons per base.

Table 10: Estimated *S. erythraea* RNA monomeric composition

Base	Frequency (%)
Adenine	22.0
Guanine	25.2
Cytosine	34.0
Uracil	18.8

3.3.2.3 Protein

Accurate analytical determination of the amino acid composition of protein is difficult requiring specialised equipment. Rather than pursue this method a simpler method based on genomic data and similar to that used to determine RNA composition was used. A closely related organism, *S. coelicolor,* was chosen. The amino acid composition of the open reading frames in the *S. coelicolor* genome was determined as stated in the method section. The composition found is given in Table 11.

Table 11: Estimated amino acid composition of *S. erythraea* protein

Amino Acid	Frequency (%)
Ala	13.7
Arg	8.4
Asn	1.7
Asp	6.1
Cys	0.8
Glu	5.7
Gln	2.6
Gly	9.7
His	2.3
Ile	2.9
Leu	10.2
Lys	2.1
Met	1.6
Phe	2.7
Pro	6.2
Ser	5.0
Thr	6.2
Trp	1.5
Tyr	2.1
Val	8.7

These results show similarity to the incomplete amino acid composition found for *S. coelicolor* by Davidson (1992). The average RMM per amino acid in protein was calculated as 106.8 Daltons per amino acid.

Based on the analysis of 16S rRNA given above *M. tuberculosis* is a closer relative of *S. erythraea* than *S. coelicolor*, however it does not contain as much genetic information as *S. erythraea.* Whilst the essential subset of the genome seems to be conserved

between actinomycetes (see Bentley *et al.* 2002) a large proportion of the genome of organisms such as *S. coelicolor* and *S. erythraea* is made up of genes encoding nonessential proteins. This means that the genes encoding rRNA and tRNA which are essential are likely to be similar between *M. tuberculosis* and *S. erythraea.* However if there is difference in amino acid composition in the gene products of the nonessential material this will not be accurately reflected by the *M. tuberculosis* genome. For this reason *S. coelicolor* open reading frames were used for estimating the *S. erythraea* amino acid composition.

3.3.2.4 Arabinogalactan

The monomeric composition of arabinogalactan was determined analytically. The ratio varied somewhat between fermentations of the same strain and varied considerably between strains. The proportions were determined independently for each fermentation. For the wild type organism between 8 and 12 arabinose monomers per galactose were found. This proportion is as expected because the galactose units form a chain with long side chains of arabinose branching off every other galactose. The ratio is similar to the proportions seen in *M. tuberculosis* (see Crick *et al.* 2001).

3.3.2.5 Lipid

The lipid composition has two layers of complexity. Firstly there are several different types of lipid in the biomass and secondly there are a large number of fatty acids available to be incorporated into these lipids. For simplicity it has been assumed here that these two levels of complexity are independent, i.e. the average fatty acid composition is independent of the lipid type.

Phospholipid composition was based on a semi quantitative analysis for *Pseudonocardia themophila* (Lechevalier *et al.* 1977) a closely related organism. Neutral lipid was assumed to be entirely composed of triacyl-glycerols. The major glycolipid found in *Saccharopolyspora* species contains 2 mannose units and a glycerol unit (Gamian *et al.* 1996). For simplicity in the model this was treated as 2.5 glucose units.

The quantity of neutral lipid at the end of the fermentation was estimated by comparing the amount of lipid at the end of the exponential phase with the amount of lipid at the end of the stationary phase. It was assumed that lipid formed in the exponential phase was phospholipid involved in the cell membrane. Lipid formed in the stationary phase was assumed to be neutral lipid accumulating as a carbon storage compound. Roughly twice as much lipid was formed in the stationary phase as in the exponential phase. To simplify the modelling this 1:2 proportion was assumed to hold true throughout the fermentation. Proportions of different lipid types used for modelling can be found in Table 12.

Table 12: Estimated composition of lipid in *S. erythraea*

Lipid type	Frequency (%)
Triacylglycerol	65.6
Phosphatidyl inositol	4.6
Phosphatidyl choline	5.1
Phosphatidyl methyl ethanolamine	5.2
Diphosphatidyl glycerol	2.7
Glycolipid	13
Phosphatidyl ethanolamine	5.3

The fatty acid composition used is the fatty acid composition of the glycolipid fraction of *S. erythraea* (ATCC 11635^T) lipid (Gamian *et al.* 1996). It is assumed that the fatty acid composition of the glycolipids is the same as for all other lipids. It is likely however that the fatty acid pool in the cell is quite variable depending on the conditions the cell experiences. Wallace *et al.* (1995) determined a rather different fatty acid composition for *S. erythraea* (ATCC 11635). Both compositions are listed in Table 13. Whilst there are differences in the chain length found they both agree that iso and anteiso fatty acid are much more common than straight chain fatty acids.

Table 13: Fatty acid composition of *S. erythraea*

Fatty Acid	Frequency (%) Gamain *et al.*	Frequency (%) Wallace *et al.*
14:0	0	0.3
14:0i	0	7.9
15:0	0	1.3
15:0i	0	14.0
15:0ai	0	22.3
16:0i	17.15	29.8
16:0	2.33	4.8
17:0i	24.22	6.8
17:0ai	48.97	11.4
17:0	4.09	1.3
18:0i	0.74	0
18:0	2.5	0

3.3.3 Biomass Composition

Once the appropriate techniques had been developed a trial fermentation was performed to establish whether biomass information should be taken into account when modelling *S. erythraea*. To this end all quantifiable biomass components, uptake rates and secretion rates were measured in batch culture. Modelling was then performed with and without the incorporation of the biomass composition data.

3.3.3.1 Fermentation

Fermentation conditions were identical to those recorded earlier in the chapter (section 3.2). The fermentation profiles are presented in three figures. Figure 11 shows the basic fermentation profile. The OUR and CER profiles show a sharp change at 40 hours. This indicates the switch from growth phase to stationary phase. The culture is nitrate limited and switches to secondary metabolism when the nitrate runs out as indicated by the start of erythromycin synthesis (see Figure 12). This phase comes to an end when the glucose is depleted after about 75 hours. During this second phase α-ketoglutarate is produced. A third phase ensues when the glucose is depleted during which α-ketoglutarate utilisation is seen. The DCW increases exponentially in the first phase, continues to slowly increase in the second phase and halts in the third phase. The OUR and CER show exponentially increasing activity in the first phase, continued

high activity in the second phase and low activity in the third phase with the transitions between phases showing drastic changes in activity.

Figure 12 contains further secretion profiles and the DNA profile. The DNA profile will be considered with the other biomass component profiles in Figure 13. The pyruvate and fumarate profiles are similar to the α-ketoglutarate profile. They are also produced in the second phase, but at much lower levels and both are reconsumed earlier. Extracellular protein is synthesised in the first phase and no further accumulation is seen in the second and third phases. Erythromycin synthesis is almost exclusively seen in the second phase.

Figure 13 contains the biomass component profiles except for the DNA profile which shown in Figure 12. The DNA, RNA, protein arabinogalactan and lipid profiles seem to show the same shape as the DCW in the growth phase. This is as expected, they are all essential integral parts of the biomass which can be expected to be synthesised proportionally during ideal growth conditions. In the second phase however this pattern changes. The DNA and the arabinogalactan continue to follow the DCW profile. The other profiles do not. Protein and RNA both contain large amounts of the limiting nutrient nitrogen. They are also both required in large amounts for rapid growth. In the second phase there is no nitrogen or rapid growth and both the RNA and the protein levels fall. Carbon storage compounds such as trehalose and lipid on the other hand accumulate proportionally faster than the DCW during the second phase. They are however reconsumed during the third phase. This basic pattern is seen in all later experiments. The details of these phases are investigated in chapter 4 and will not be further explored here.

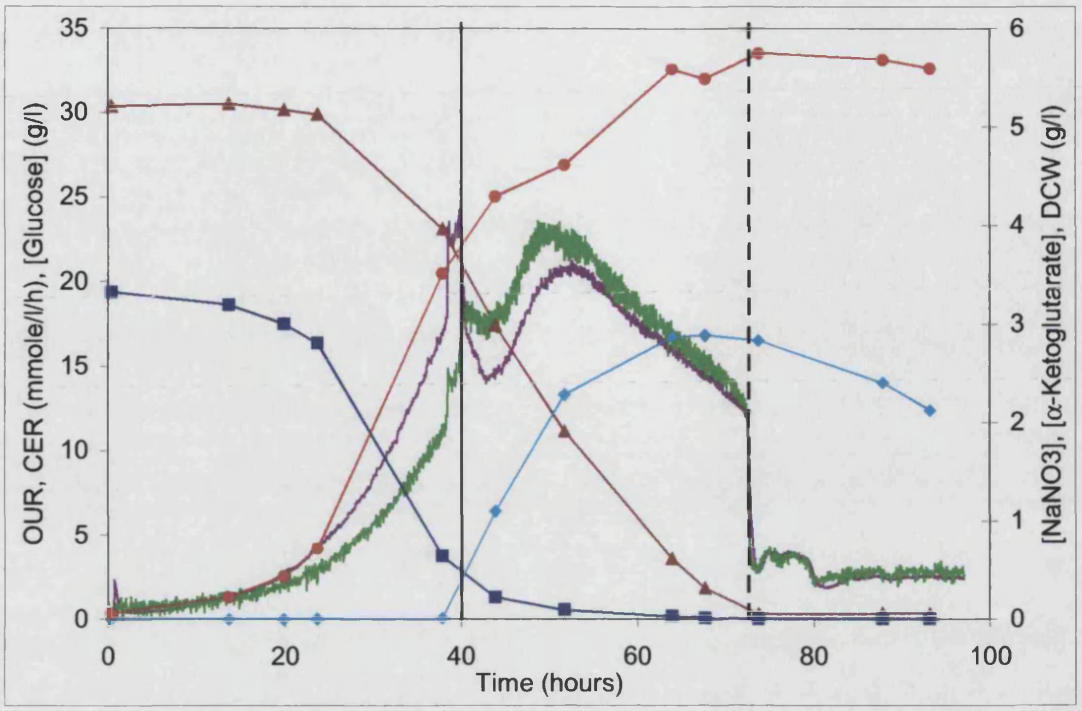


Figure 11: White variant wild type fermentation 3, profile 1. OUR (mmole/l/h) —, CER (mmole/l/h) —, Glucose (g/l) ▲, Nitrate (g/l) ■, α-Ketoglutarate (g/l) ◆, Dry cell weight (g/l) ●. Dashed lines mark the transition from growth phase to stationary phase and the depletion of glucose from left to right respectively.

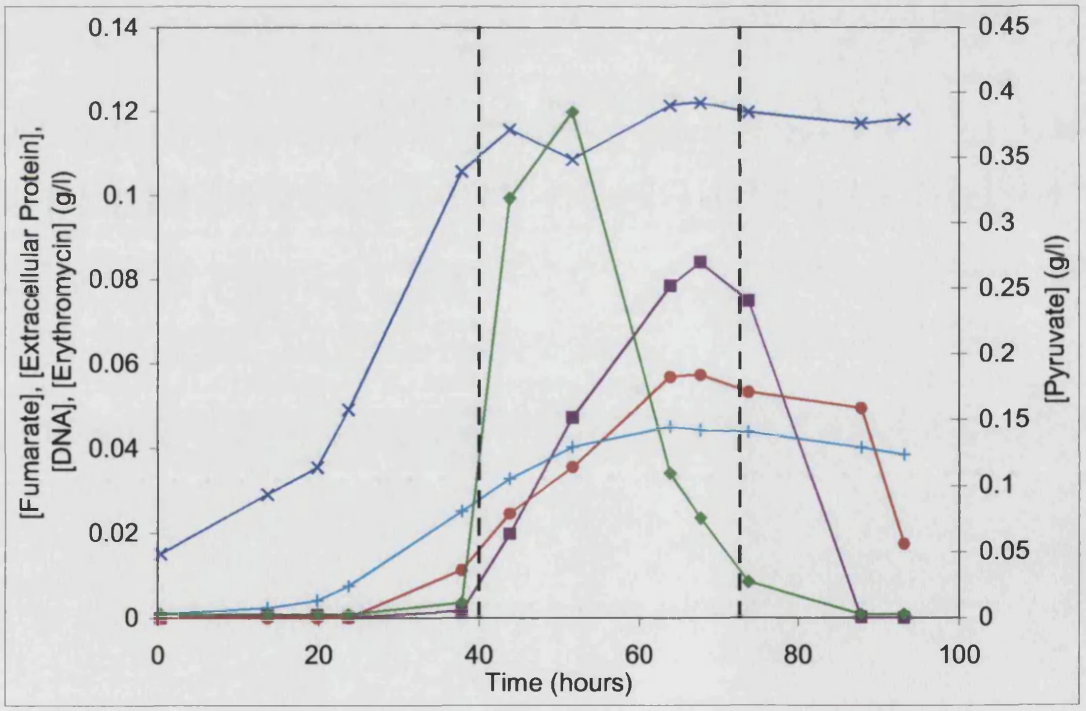


Figure 12: White variant wild type fermentation 3, profile 2. Fumarate (g/l) ■, Extra cellular protein (g/l) ×, Erythromycin (g/l) ●, DNA (g/l) +, Pyruvate (g/l) ◆. Dashed lines mark the transition from growth phase to stationary phase and the depletion of glucose from left to right respectively.

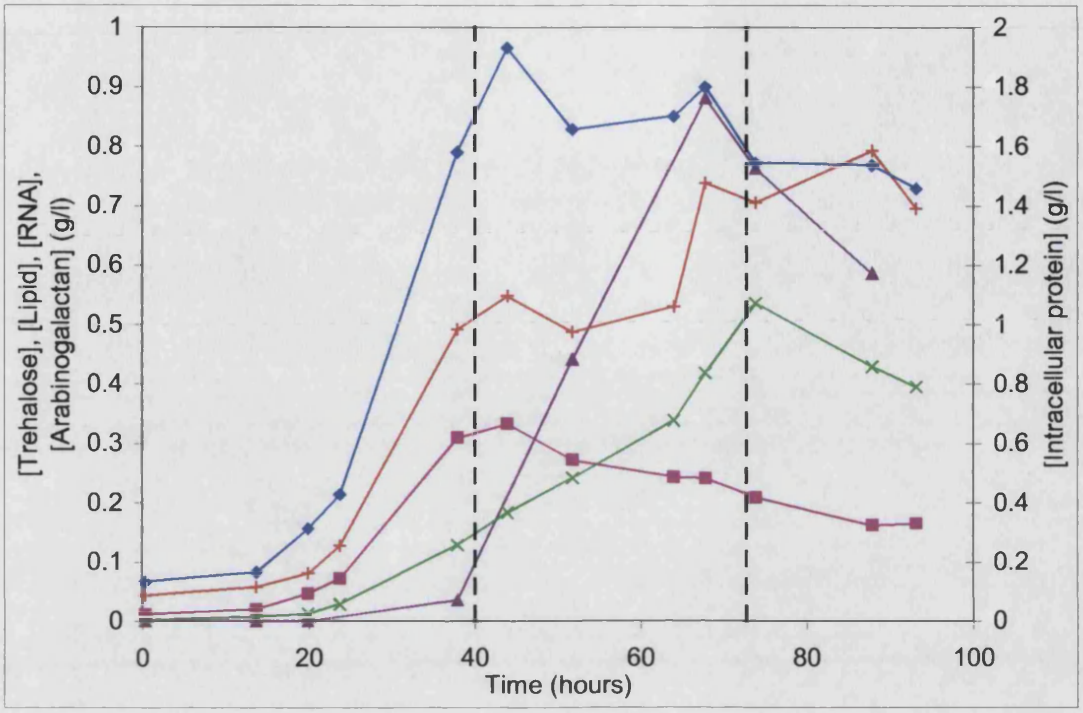


Figure 13: White variant wild type fermentation 3, profile 3. RNA (g/l) ■, Lipid (g/l) ×, Trehalose (g/l) ▲, Arabinogalactan (g/l) +, Intracellular protein (g/l) ◆. Dashed lines mark the transition from growth phase to stationary phase and the depletion of glucose from left to right respectively.

3.3.3.2 Biomass Composition

The changing levels of the biomass components relative to the dry cell weight are summarised in Figure 14. Figure 15 contains the equivalent figure from white variant wild type fermentations 4 in chapter 4 for comparison. The profile for white variant wild type fermentation 4 contains a trehalose profile obtained with a more sensitive instrument than that used in this chapter. Trehalose can be seen to be present at low levels during the growth phase using this more sensitive instrument. Otherwise the profiles are broadly similar.

The biomass components measured account for around 80% of the biomass during the first phase of the fermentation. This figure slowly drops during the second and third phases to around 60%. A discrepancy between the DCW and the total mass of the measured biomass components was expected since not all components were measured. For example the peptidoglycan in the cell walls was not measured. For the closely related *S. coelicolor* Davidson (1992) found that this made up 4.6% of the DCW.

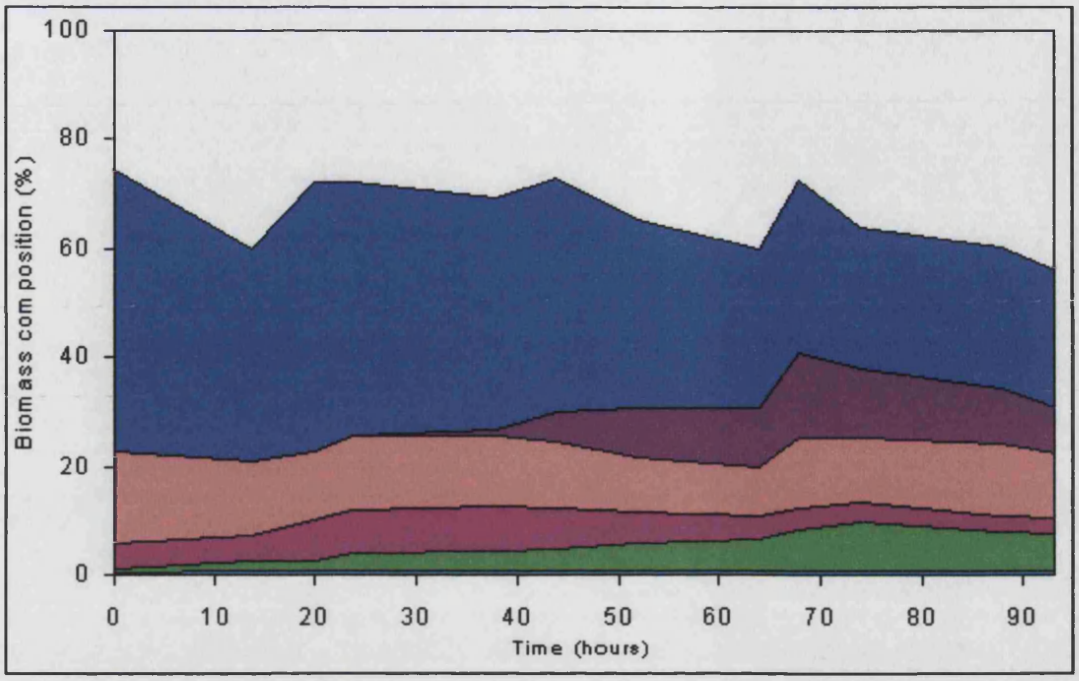


Figure 14: Biomass composition for white variant wild type fermentation 3.

Protein, trehalose, arabinogalactan, RNA, lipid, DNA, Unknown.

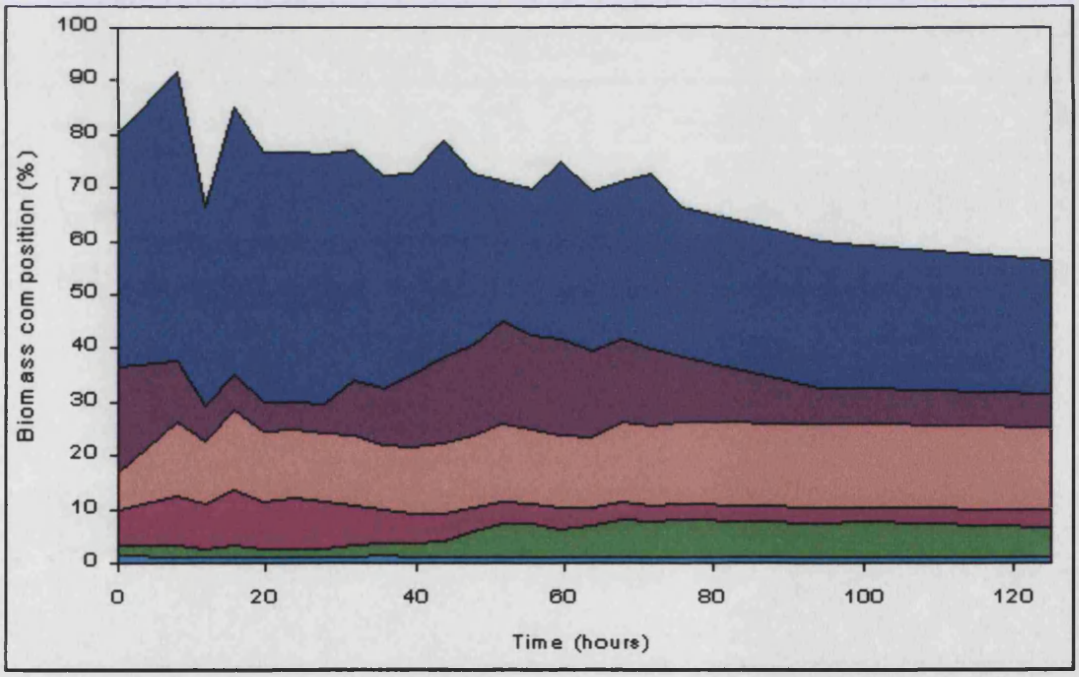


Figure 15: Biomass composition for White variant wild type fermentation 4. In this profile a more sensitive instrument was used to measure trehalose than in the profile above.

Protein, trehalose, arabinogalactan, RNA, lipid, DNA, Unknown.

The cell also contains metabolic intermediates the sum total of which is known as the metabolic pool. The size of this is difficult to estimate. Neidhardt *et al.* (1987) estimated that the pool composed 2.5% of the DCW of *E. coli*. However Pai and Kubitscheck (1992) looked at a much wider range of intracellular metabolites and concluded an average of 8-9% which could double at certain stages in the cell cycle. Mousdale (1998) notes that "the cellular pools of low molecular weight metabolites are much greater in prokaryotic and eukaryotic organisms exhibiting secondary metabolism than in laboratory enteric bacteria such as *Escherichia coli*".

Another source of error is residual water. The drying method used to take DCW readings usually leaves a small amount of residual water bound tightly to the biomass. Roubos (2002) estimated this to be 4% of the DCW in *S. clavuligerus*.

The biomass also contains inorganic salts which add to the unmeasured component. Roubos (2002) found 5% of the biomass was composed of sulphate and 4% of phosphate in *S. clavuligerus*. Duboc *et al.* (1995) measured sulphate and phosphate for *S. erythraea* and found the biomass contained 0.84% sulphate and 3.35% phosphate. They also estimated the ash content which includes these minerals as 14.5%.

Given the number of unmeasured components 80% represents a reasonable proportion of the biomass for the measured components. For comparison Davidson (1992) accounted for 82% of the DCW in *S. coelicolor*. Roubos (2002) accounted for 60-70% of the DCW in *S. clavuligerus*.

The fall in the proportion of the biomass accounted for in the measurements towards the end of the fermentation is probably due to the accumulation of an unmeasured component. This fits well with the observation that the nitrogen balance does not close in the stationary phase suggesting that a nitrogen containing compound is being accumulated in the biomass. Section 3.3.4 explores the possibility that this is due to the formation of a capsule. The closely related *M. tuberculosis* produces a polyglutamate capsule. This could account for both the observed anomalies in the data.

Protein makes up the largest measured component of the biomass. It generally forms just over 40% of the biomass during the first phase of the fermentation. This however

falls to around 25-30 % in the stationary phase. Likewise the RNA which makes up around 8% of the biomass in the exponential phase drops to around 3% in the stationary phase. The fate of the nitrogen released by the degradation of protein and RNA is unknown but it may be used in the formation of a polyglutamate capsule (see section 3.3.4). These profiles are typical of nitrogen limited actinomycetes. King and Büdenbender (1997) found very similar profiles for both protein and RNA with 6 *streptomyces* strains. Doskočil *et al.* (1959) tried to estimate DNA and DNA synthesis in *Streptomyces aureofaciens* under nitrogen limitation by measuring the consumption of inorganic phosphate. Comparison with the results found here is difficult however as the total amount of biomass was not reported. Their results are in any case likely to be spurious as *streptomyces* form teichoic acids which contain phosphorous.

Trehalose accumulates in the biomass during the stationary phase. There seems to be a difference in the trehalose profiles between the figures with Figure 14 showing trehalose present only in the stationary phase whereas Figure 15 shows low levels of trehalose present in the growth phase. The latter figure is probably the truer reflection of the biomass composition as the data for this profile was collected with a more sensitive detector. All the profiles show significant accumulation of trehalose in the stationary phase. To my knowledge this is the first time trehalose has been quantified in *S. erythraea.* Trehalose has been noted as present in actinomycetes previously for example by Braña *et al.* 1986. They considered that its accumulation pattern suggested that its role was in protection from desiccation rather than storage of carbon and energy, however they did not rule these out. The use of trehalose as a storage compound has been reported before. Elbein *et al.* (2003) reported that trehalose is used as an energy and carbon reserve in fungal spores which hydrolyse it on germination. The accumulation of trehalose in the stationary phase probably has more than just a carbon storage role. Elbein *et al.* (2003) reported seven roles of trehalose, a number of them were related to its protective properties. Trehalose could play an important role in protecting aerial mycelia and spores, and even though these are not seen in submerged culture the build up of trehalose may still occur. Schneider *et al.* (2000) suggested that in *Streptomyces coelicolor* trehalose might be used as a means to carry carbon from storage in glycogen to the places that it is needed within a colony. However given the failure to detect glycogen in *S. erythraea* this role seems unlikely in this case.

Lipids are seen throughout the fermentation profile. Through the growth phase they make up around 3% of the DCW and this probably represents polar lipids in the cell membrane. Lipid also accumulates in the stationary phase until it makes up around 9% of the dry cell weight. This accumulation is probably in the form of neutral lipids which are storing the excess carbon available to the cell when the nitrogen source runs out. The levels of lipid are a little lower than expected. Stouthamer (1979) found that *E. coli* contained 9% lipid, this is equivalent to the growth phase result as *E. coli* does not display secondary metabolism. Nielsen (1997) found 5% lipid in *Penicillium chrysogenum.* In more closely related species Davidson (1992) found an average value over the whole fermentation for *S. coelicolor* of 5.5%. This is quite similar to the proportions found here if averaged over the fermentation. Roubos (2002) found lipid to be 7-14% of the DCW of *S. clavuligerus* during the growth phase. The difference between the lipid levels found by Roubos and those found in this study and in the Davidson thesis may be due to the method used. Roubos used a gravimetric method in which the hydrophobic biomass constituents are extracted into organic solvent and dried and weighted. This is likely to over estimate the lipids as other hydrophobic substances may also be weighted. By contrast the vanillin method used in this work is likely to underestimate the lipid content as different lipids give different signal strengths. Triolene was used as the standard this gives a stronger signal than polar lipids leading to an under estimation of their abundance.

The DNA levels stay reasonably constant throughout the fermentation at about 1% of the DCW. This is a little higher than Roubos' 0.4-0.6% in *S. clavuligerus* and somewhat lower than Davidson's 6% for *S. coelicolor* and Shahab *et al.* (1996) who found 3.6% for *S. coelicolor.* The findings regarding biomass composition found in this work and those found in the literature are summarised in Table 14.

Although *S. erythraea* shows a similar patterns of morphology and colony development to the *streptomyces* the biomass composition of *S. erythraea* has some important differences in cell wall composition. In *S. erythraea* no teichoic or lipoteichoic acids are formed. Its cell wall composition is more similar to that of the *mycobacteria* containing arabinogalactan and producing a capsule. Unlike the *mycobacteria* however it produces no mycolic acids.

Table 14: Biomass composition of *S. erythraea* and some related organisms

Component	*S. erythraea*	*S. erythraea*	*S. clavuligerus*	*S. coelicolor*	*S. coelicolor*
Mode of growth	Batch culture	Batch culture	Batch culture	Batch culture	Continuous culture
Phase	Exponential	Stationary	Exponential	Average	$\mu = 0.109$
Reference	This work	This work	Roubos[§] (2002)	Davidson (1992)	Shahab *et al.* (1996)
Protein	42%	27%	44-45%	53%	41%
AG	16%	12%	0-6%*	nd	nd
Trehalose	5%	13%		nd	nd
RNA	8.6%	3.6%	11-15%	14%	17%
Lipid	3.1%	9.2%	7-14%	5.5%	nd
DNA	1.0%	0.76%	0.5-0.7%	6%	3.6%
Murein	nd	nd	8% est	4.6%	nd
Unknown	25%	34%	23-42%	17%	38%

[§] Several biomass compositions were reported obtained under different growth conditions. This composition was selected because the carbon source used (maltose) is broken down to glucose, the carbon source used in this work.

* total carbohydrate measured.

AG denotes arabinogalactan

nd denotes that the measurement was not determined in this case.

est denotes that the murein content reported by Roubos is estimated based on the measured components and the elemental composition of the biomass.

3.3.3.3 Summary

The biomass composition of *S. erythraea* does change significantly during the course of a nitrogen limited fermentation. This change is primarily characterised by: 1) diminishing levels of the two main nitrogen containing compounds (protein and RNA) during the stationary phase and 2) accumulation of carbon storage compounds (trehalose and lipid). There is also a large amount of biomass unaccounted for. This is probably made up of unmeasured components such as murein, salts, metabolite pools and tightly bound water in the biomass. In the stationary phase this unknown portion increases, this is thought to be due to the accumulation of an unknown nitrogen containing biomass component possibly capsular material.

3.3.4 Capsule Formation

3.3.4.1 Introduction

Some bacteria produce capsules of polymerised material that surrounds the cell in a slimy coat. Capsules of polysaccharide (sugar derived) and polyglutamate (amino acid derived) are seen in a number of Gram positive bacteria, eg *Bacillus anthracis* (Mesnage *et al.* 1998) and *Mycobacterium tuberculosis* (Harth *et al.* 1999). Both produce polyglutamate capsules.

A number of strands of evidence point to the synthesis of such a capsule in *S. erythraea* during the stationary phase. *S. erythraea* displays extensive wall growth in the stationary phase. Moreover the pipetting properties of the broth change becoming thicker during the stationary phase suggesting a change in the viscosity of the broth. In addition the nitrogen balance in the stationary phase does not close. Although large amounts of protein and RNA are broken down there is no significant increase in nitrogen containing biomass components or extracellular protein. Annotation work on the *S. erythraea* genome has revealed that the organism contains a gene for polyglutamate formation (P. Krabben, personal communication). The formation of a polyglutamate capsule during the stationary phase could explain all these observations.

Capsules are made by a number of pathogenic bacteria and simple dye based procedures such as the India ink method have been developed to identify them. Quantifying and analysing the composition of the capsule are however more involved. In order to try to explain the unusual observations noted above the India ink method was used to test for capsule formation.

3.3.4.2 Materials and Methods.

3.3.4.2.1 Growth and image acquisition

S. erythraea was grown in 50 ml of nitrogen limited media (see section 2.1) in 500 ml shake flasks (10% working volume). Samples were taken in mid exponential phase and

mid stationary phase stained and studied under a Leica DMRA2 microscope. Images were digitally acquired using a CCD camera which captured greyscale images of 760 by 574 pixels (Leica Microsystems UK Ltd, Milton Keynes, Beds, U.K.). No digital enhancement has been performed on the images. All images are presented at the same level of magnification.

3.3.4.2.2 Staining

India ink is a slightly acidic dye and cannot penetrate the capsular material. Cells stained with this dye show a halo or clear area around the cell if capsular material is present. In order to make the distinction between the cell and the capsule more obvious cells can be counter stained with an alkaline solution of methylene blue (Loeffler's reagent), which stains the cells but not the capsule. The cells can then be seen as bright blue surrounded by a clear capsule surrounded by a dark background (see http://www.arches.uga.edu/~jarneson/capsulestain.html).

India ink was purchased from a local art shop and was manufactured by Winsor and Newton. It was used without modification. Methylene blue (general purpose grade) was purchased from Fisher Scientific. Methylene blue (0.5g) was dissolved in distilled water (100 ml) at 50°C. To this 1 ml of 1% KOH solution and 30 ml of 95% ethanol were added.

One drop of cells was mixed with one drop of India ink and spread thinly across the slide with a cover slip. The slide was left to thoroughly air dry. The slide was then gently flooded with Loeffler's reagent and left to fix for 3 minutes. The Loeffler's reagent was drained from the slide and the slide was gently washed with distilled water to remove the excess stain without disturbing the cells. The slide was thoroughly air dried. Immersion oil was applied directly to the slide immediately before viewing.

3.3.4.3 Results and Discussion

Seventy six photographs were taken of cells at various points in the exponential and stationary phase. For brevity representative examples of the observed structures are

shown. The photos are presented in black and white as Figure 16 and Figure 17. Structures stained blue are marked as blue, however the visual impact is reduced in monochrome. All photos are at the same magnification.

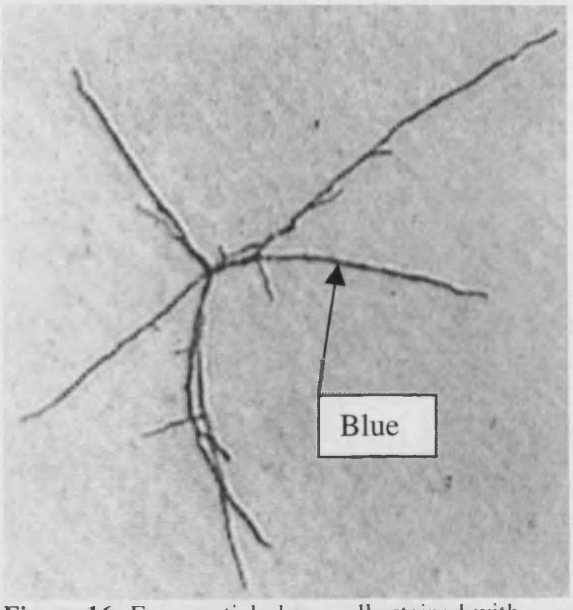


Figure 16: Exponential phase cells stained with India ink and Loeffler's methylene blue stain. Note the cells stain blue and appear dark.

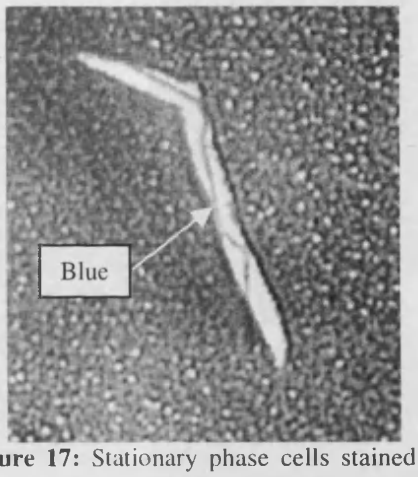


Figure 17: Stationary phase cells stained with India ink and Loeffler's methylene blue stain. Note the cells stain blue and appear dark, the capsule repels the dark India ink and remains light.

Cells harvested in the exponential phase show no capsule. When stained with both India ink and Loeffler's reagent no clear area is visible and the cells appear an intense blue colour (see Figure 16). Cells harvested in the stationary phase appear larger and show a considerable capsule. When these cells are also stained with Loeffler's reagent the mycelia become visible in the capsule as an intense blue vein in the clear surrounding material (see Figure 17).

The formation of a capsule represents a biomass component that has not been accounted for in the composition measured in section 3.3.3.2. This explains why the percentage of biomass accounted for is lower in the stationary phase than in the exponential phase. Harth and Horwitz (1999) reported that polyglutamate in *M. tuberculosis* made up about 50 μg per 10^{10} cells. Assuming a cell mass of 2.9×10^{-13} g/cell wet weight and dry cell weight is one quarter of wet cell weight polyglutamate makes up approximately 7% of the dry cell weight of *M. tuberculosis*. If the levels are similar in *S. erythraea* this would have a large adverse effect on the nitrogen balance. Polyglutamate probably has a second unfortunate effect on biomass measurement as it may be the cause of the wall growth seen which removes biomass from the bulk media reducing the measured

biomass concentration. Wall growth causes the effected biomass to experience conditions considerably different to those in the bulk media. Flux balance analysis relies on the biomass experiencing uniform conditions leading to uniform metabolic behaviour.

S. erythraea exhibits morphological differentiation when grown on solid media (a more natural substrate than liquid culture). This involves the formation of aerial mycelia. For this reason it is not surprising that cell surface changes as the cells move from exponential phase to stationary phase. The formation of a capsule seen here is probably part of this process which of course cannot be completed in submerged culture. The formation of wall growth does however hint at cells coming together in a way that they don't in the exponential phase.

Time did not permit the nature and quantity of the capsular material to be investigated. However given that: 1) The nitrogen balance does not close in the stationary phase. 2) The closely related *M. tuberculosis* exhibits a polyglutamate capsule. 3) One of the genes used in the synthesis of a polyglutamate capsule has now been identified in *S. erythraea* (P. Krabben personal communication). It seems likely that the capsule is formed of polyglutamate.

3.3.4.4 Conclusions

S. erythraea does not exhibit a capsule during the exponential phase but develops one when the cells enter stationary phase. While the nature of this capsule is not known for certain, circumstantial evidence points to it being made of polyglutamate. The formation of this capsule could be one of the causes of the shortfall in mass of the biomass components compared to measured dry cell weight seen in the stationary phase. It may also account for changes in the rheology of the broth and for wall growth in the stationary phase.

This work could be usefully extended by analytically determining the quantity and composition of the capsular material. This information should help improve the accuracy of the flux distributions found by flux analysis.

When the complete genome is available for *S. erythraea* it may be possible to knockout the genes for polyglutamate formation in order to verify our findings. This might also be useful for preventing wall growth and improving process performance.

3.3.5 Comparison of Two Methods For Flux Balance Analysis

Section 3.3.3.2 has shown that large changes occur in the biomass composition of *S. erythraea* during the stationary phase. In order to assess the impact of this on modelling flux balance analysis was performed twice, once with biomass composition data and once assuming a constant biomass composition. For clarity the calculations performed with measured biomass composition will be referred to as the dynamic model and the calculations performed assuming a constant biomass composition will be referred to as the static model.

Variance weighted partial least squares analysis (see section 2.8) was used for this analysis as it is a powerful method for minimising errors in measured data. If significant differences between the flux distributions were found using a method designed to minimise these errors then there would be a very strong case for measuring the biomass composition during fermentations. For the control case distributions were calculated based on measured biomass data for the relevant time during the fermentation. For the test case the average biomass composition in the exponential phase was used for all time points. This should lead to similar flux distributions in the growth phase, however differences should be seen as the biomass composition starts to change in the stationary phase. Daae and Ison investigated the impact of the differences of biomass composition between species. By contrast this approach tackles changes of biomass composition during a fermentation.

Time points were selected for the performance of flux balance analysis to represent the whole fermentation. Three time points were chosen during the exponential phase and three during the stationary phase, two before glucose depletion and one after glucose depletion. Weighted residuals are used to copmpare flux distributions. For an explaination od weighted residuals see section 2.8.2. Flux distributions are often presented on a common basis eg glucose uptake = 100 mmole/gDCW/h. This is not

done here because in some flux distributons the glucose uptake rate is zero and no other suitable rate could be found.

3.3.5.1 Comparison of Exponential Phase Flux Distributions

Three pairs of flux distributions have been calculated for the exponential phase. They were calculated for 16, 26 and 36 hours. Those calculated with the dynamic model are shown in Figure 18, Figure 20 and Figure 22 respectively. Those calculated with the static model are shown in Figure 19, Figure 21 and Figure 23 respectively. The weighted residual is also given. This is a measure of how closely the measured fluxes agree with the values arrived at by the fitting procedure for these fluxes (see section 2.8.2). Any measured flux found to cause a very high weighted residual was not used in the calculation, when this occurred a note is provided saying that the flux was undefined.

The group of flux distributions for the exponential phase have many features in common and will be considered together. The main feature in the exponential phase is the high degree of similarity between the distributions calculated using the two different methods. The distributions calculated at 26 hours will be used to illustrate this. The fluxes through glycolysis, and the pentose phosphate pathway (PPP) are almost all identical to within ±1%, the fluxes in the TCA cycle are within ±5%. The only really significant difference is for the flux from the PPP to biomass where there is a 15% difference. This high degree of similarity for the growth phase is as expected. The biomass composition used for the static model is the average for the growth phase where constant biomass composition is both expected and experimentally found. This means that as long as the growth rate is correctly calculated the flux distributions for the two methods should be the same.

The difference in the flux from the PPP to biomass probably relates to the synthesis of arabinose for the cell wall. The ratio of arabinose to galactose varies slightly from fermentation to fermentation and within fermentations. For the static model this ratio is fixed. When the dynamic model is used this ratio reflects the actual ratio at the time of sampling. The unusual high fluxes through the PPP will be discused in chapter 4.

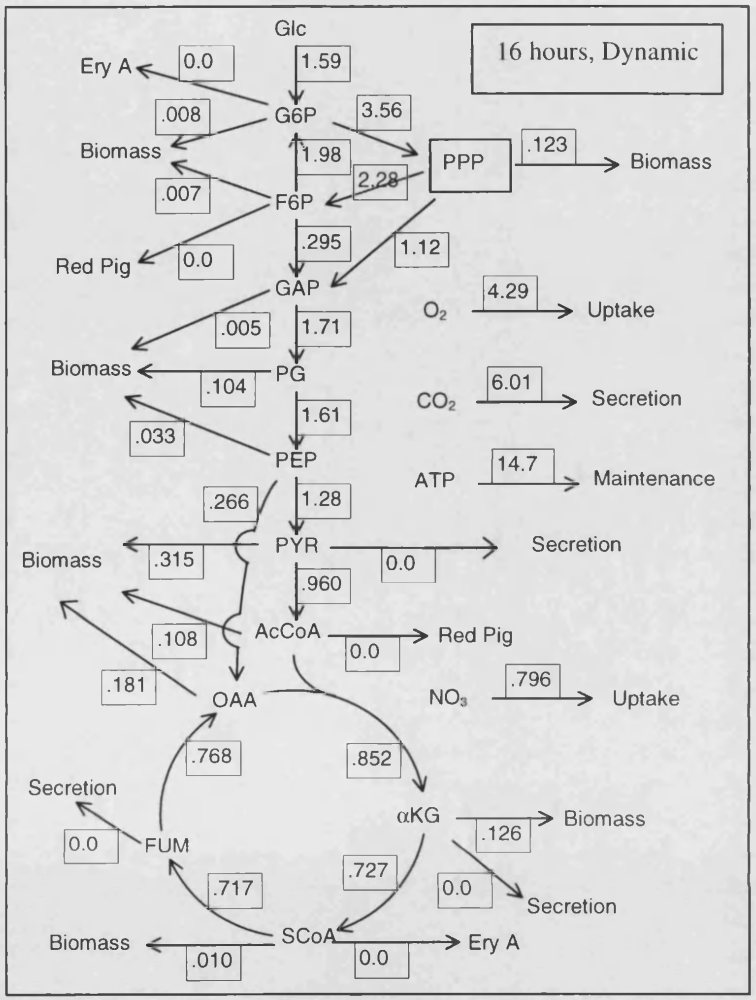


Figure 18: White variant wild type flux distribution at 16 hours based on measured data. Weighted residual = 0.89. Glucose flux undefined. All fluxes are in mmole/gDCW/hour

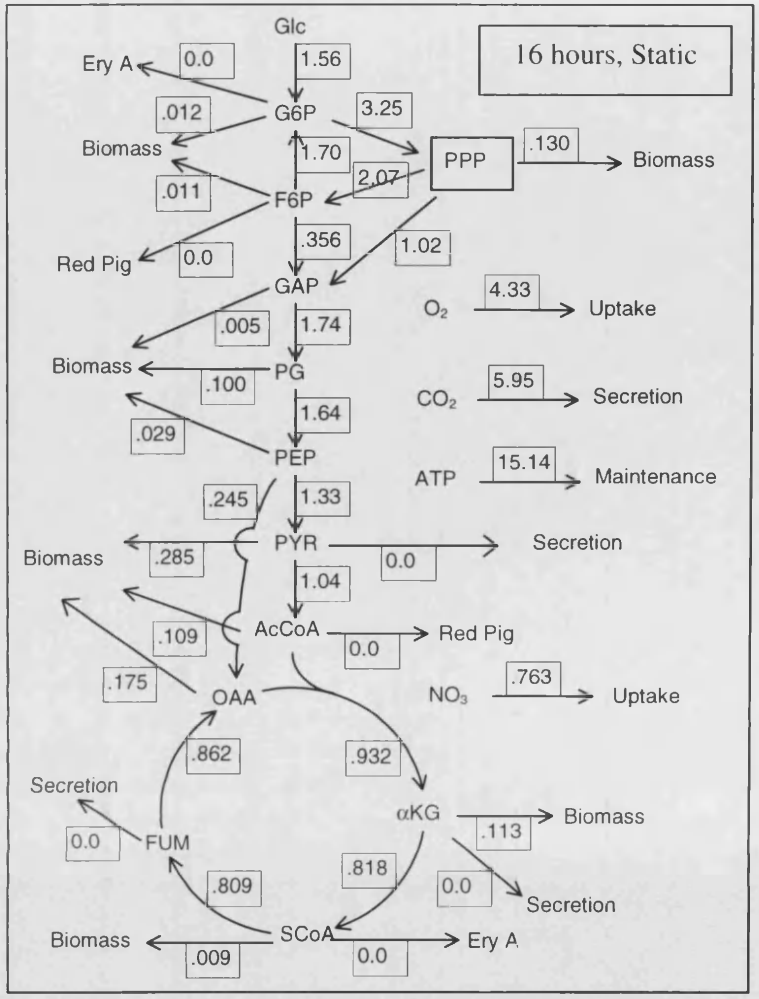


Figure 19: White variant wild type flux distribution at 16 hours based on assumed data. Weighted residual = 0.14. Glucose flux undefined. All fluxes are in mmole/gDCW/hour

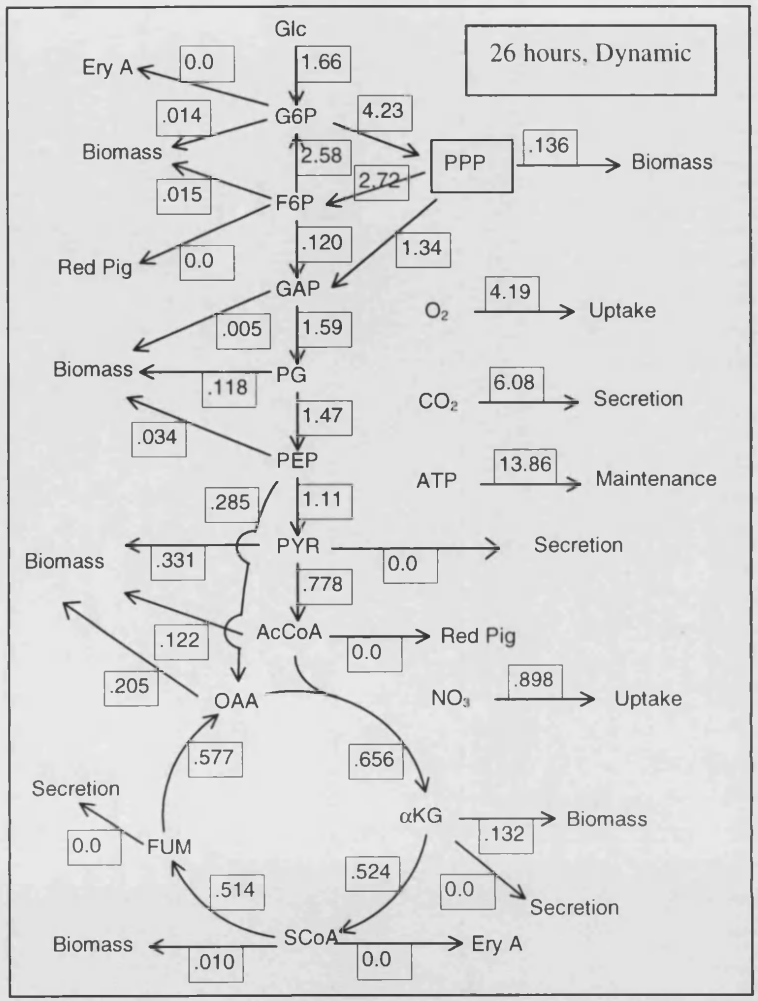


Figure 20: White variant wild type flux distribution at 26 hours based on measured data. Weighted residuals = 0.73. Glucose flux undefined. All fluxes are in mmole/gDCW/hour

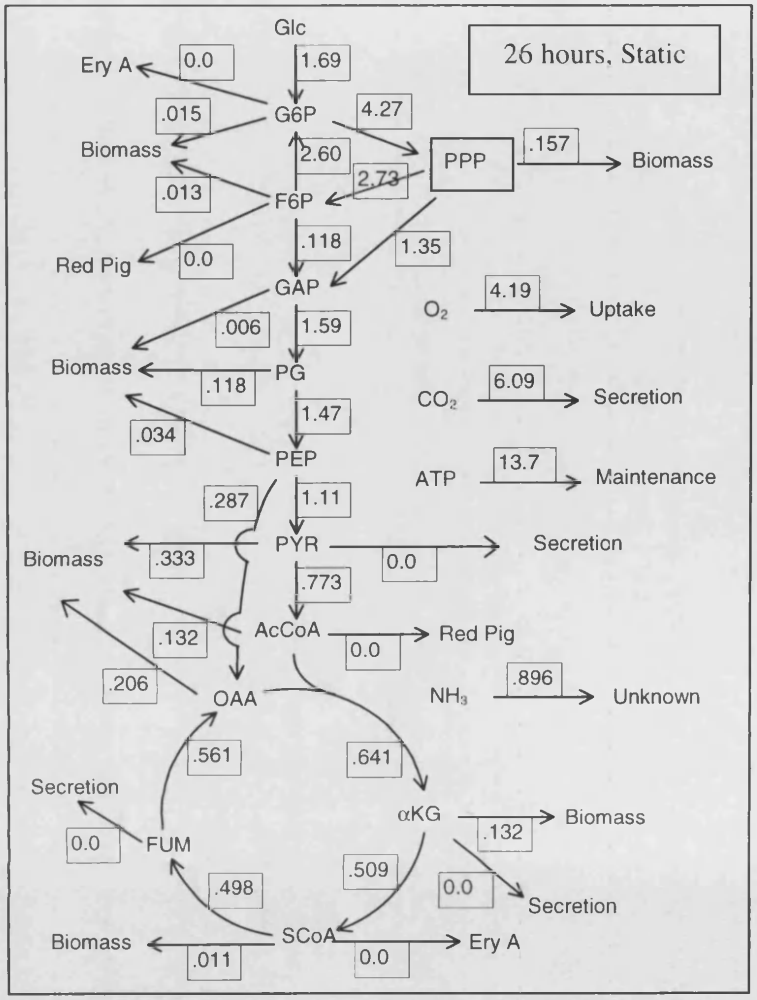


Figure 21: White variant wild type flux distribution at 26 hours based on assumed data. Weighted residual = 1.2. Glucose flux undefined. All fluxes are in mmole/gDCW/hour

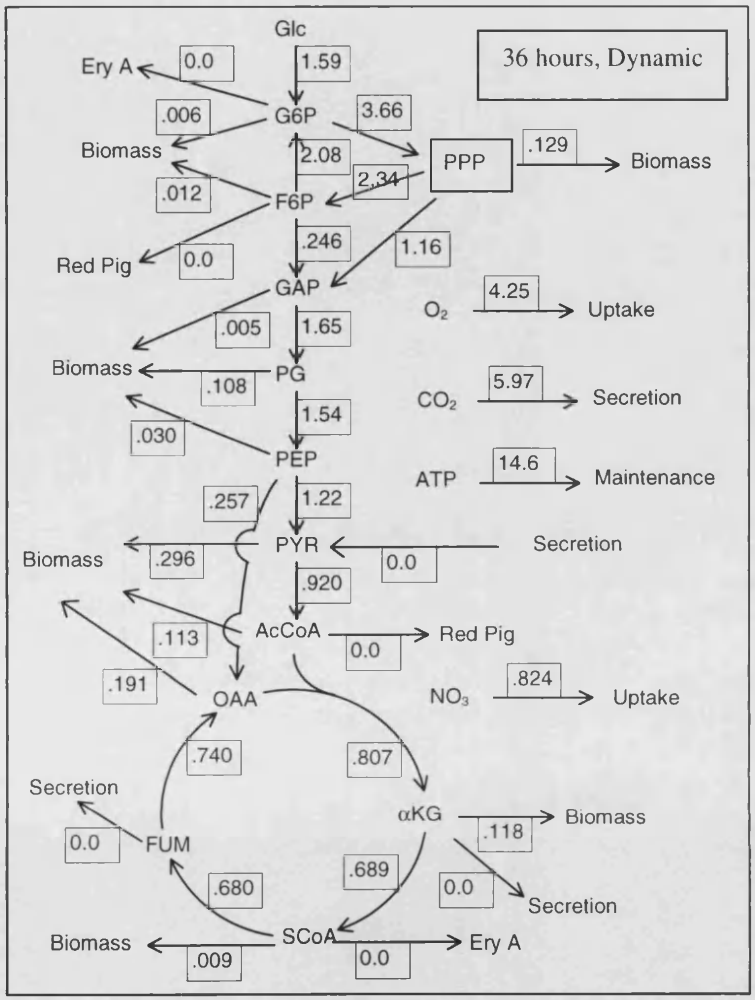


Figure 22: White variant wild type flux distribution at 36 hours based on measured data. Weighted residuals = 4.0. Nitrate flux undefined. All fluxes are in mmole/gDCW/hour

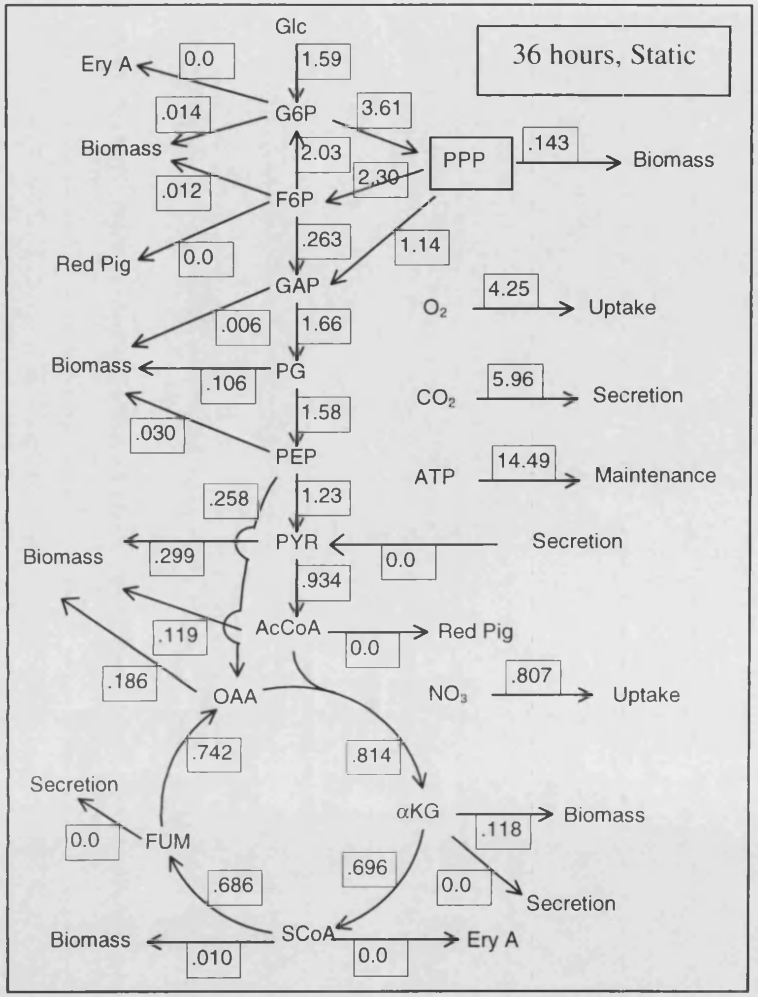


Figure 23: White variant wild type flux distribution at 36 hours based on assumed data. Weighted residual = 4.4. Nitrate flux undefined. All fluxes are in mmole/gDCW/hour

The weighted residual for the two methods are similar in size as would be expected with similar biomass compositions. The residual for the dynamic model is smaller than for the static model, this is also true at 36 hours. Although giving similar results the flux distribution calculated with the dynamic model is more consistent with the measured data than the distribution calculated with the static model. The magnitude of the residuals at 36 hours suggests however that the data used here is not as constant as the data used at 26 hours. The 36 hour time point is very close to the transition from growth phase to stationary phase. It could be that the growth is not as balanced at this stage as in mid exponential phase, this would adversely effect determination of the flux distribution.

The flux distributions at 26 and 36 hours show extremely close similarity between the dynamic model and the static model. The distributions for 16 hours are not quite as close. This is probably due to the low levels of biomass at this stage of the fermentation which make accurate determination of the composition more difficult. The weighted residuals suggest this is the case. The residual for the dynamic model is much higher than the residual for the static model showing inconsistencies exist in the measured data.

3.3.5.2 Comparison of Stationary Phase Flux Distributions Before Glucose Depletion

In the stationary phase the biomass composition of *S. erythraea* changes. It is during this phase that the use of the static model is expected to give incorrect flux distributions. This is because no provision is made for independent synthesis or break down of biomass components in the static model. Two pairs of flux distributions for the stationary phase before glucose depletion have been calculated one at 52 hours and one at 62 hours. The results from the dynamic model are presented as Figure 24 and Figure 26 respectively and the results from the static model are presented as Figure 25 and Figure 27 respectively.

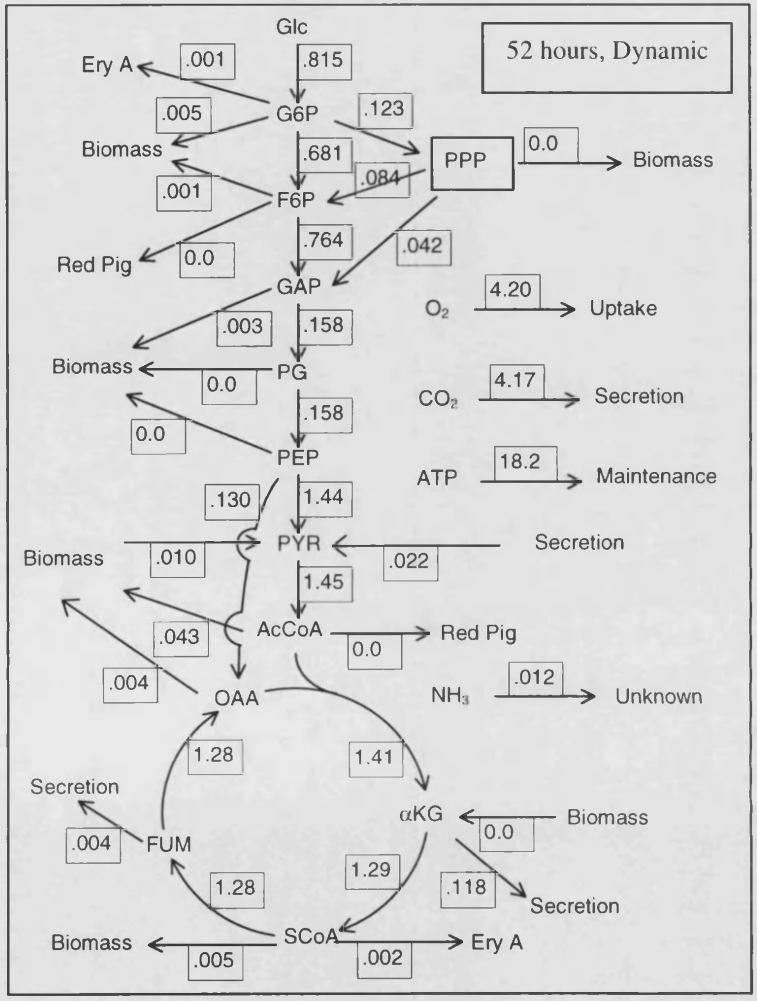


Figure 24: White variant wild type flux distribution at 52 hours based on measured data. Weighted residual = 1.1. Flux to NH_3 undefined. All fluxes are in mmole/gDCW/hour

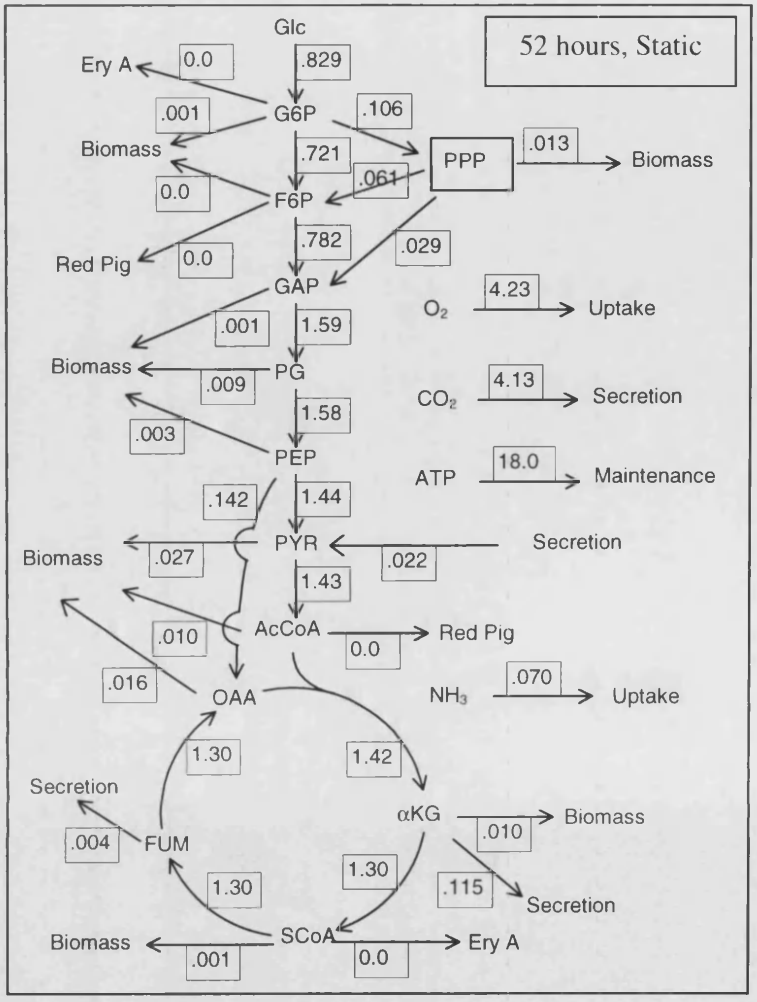


Figure 25: White variant wild type flux distribution at 52 hours based on assumed data. Weighted residual = 1.9 flux to NH_3 undefined. All fluxes are in mmole/gDCW/hour

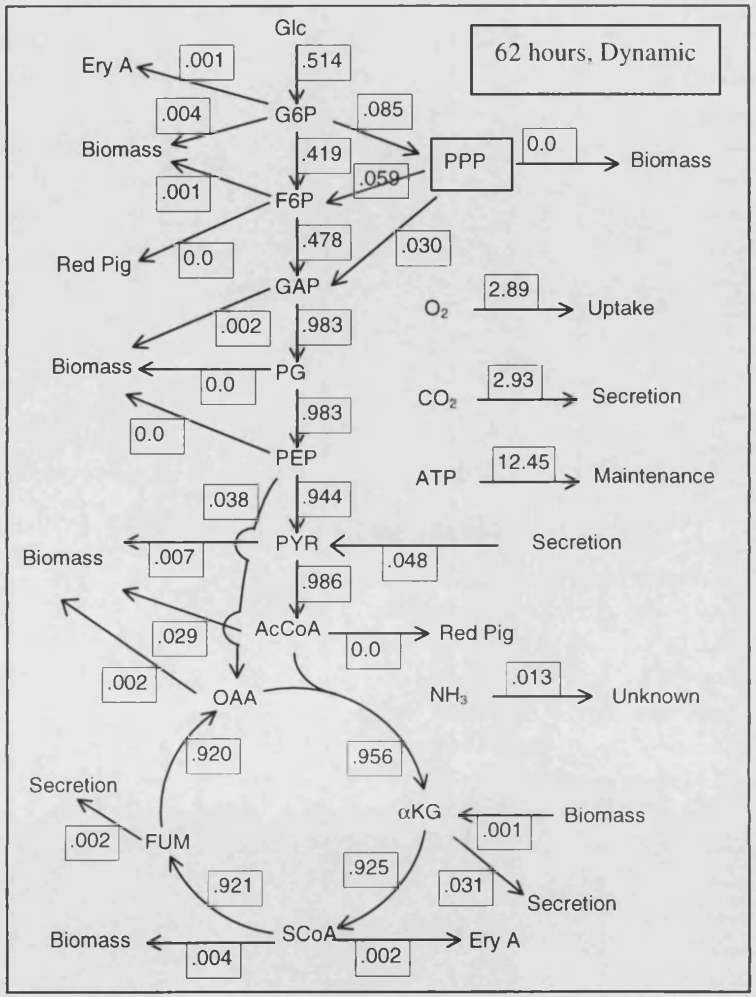


Figure 26: White variant wild type flux distribution at 62 hours based on measured data. Weighted residual = 0.40. Flux to NH_3 undefined. All fluxes are in mmole/gDCW/hour

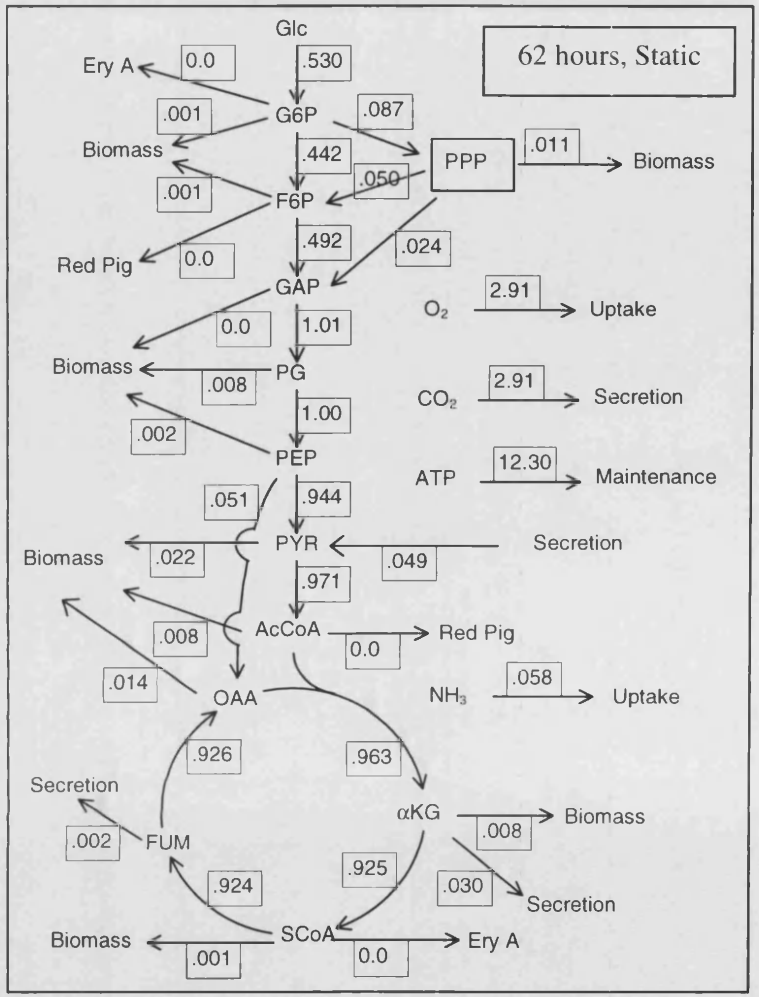


Figure 27: White variant wild type flux distribution at 62 hours based on assumed data. Weighted residual = 1.1. Flux to NH_3 undefined. All fluxes are in mmole/gDCW/hour

Note that due to the poor closure of the nitrogen balance in this and the next phase a hyperthetical flux for ammonia has to be included in the modle to maintain the steady state. This will be discussed futher in chapter 4.

The differences between the flux distributions calculated using the different methods during this phase are larger than the differences found for the exponential phase. However the differences in central metabolism are not as large as expected. The flux distribution at 52 hours will be taken as an example. The fluxes through glycolysis and through the TCA cycle are almost identical. Flux through the PPP however is different. The flux through the oxidative branch is 16% higher in the dynamic model than in the static model. The difference is greater in the non-oxidative branch because in the static model some of the flux is diverted to biomass formation lowering the fluxes in this branch.

The difference in the PPP flux illustrates one of the key differences between the two flux distributions. The amount of biomass is still slowly increasing at this stage. In the dynamic model this extra biomass is provided by the accumulation of lipid and trehalose. These fluxes are greater and there are no fluxes to amino acids and ribonucleotides. In the static model the balanced biomass production assumed at this stage leads to smaller fluxes to trehalose and lipid and larger fluxes to amino acids and ribonucleotides. The heavy demand for NADPH in lipid synthesis probably leads to the higher flux through the oxidative branch of the PPP for the dynamic model.

With the exception of the PPP the differences in the central metabolic pathways are not particularly large. The fluxes in the rest of metabolism do show some large differences. One example is the use of nitrogen. In the dynamic model the network produces excess nitrogen from the catabolism of RNA and protein. As the destination of this nitrogen is unknown it is represented as a flux from NH_3 to unknown. In the static model the opposite is seen, the organism is assumed to be still producing protein and RNA. This requires a significant uptake of nitrogen into the cell, which is not feasible since all nitrate has already been consumed. Without taking the biomass composition into account the importance of the nitrogen balance in this phase would not have been clear.

Some of the fluxes to biomass components show opposite directions between the two models. For example at the pyruvate node the flux in the dynamic model is from biomass to pyruvate. However for the static model the flux is in the opposite direction. This is because pyruvate is both an important precursor of amino acid synthesis, which occurs in the static model and an important product of amino acid catabolism, which occurs in the dynamic model.

Fluxes to the storage compounds lipid and trehalose are much higher (approximately 4-5 fold) in the dynamic model than in the static model (see fluxes of acetyl-CoA and glucose-6-phosphate to biomass). This occurs because the extra biomass synthesised in the stationary phase is assumed to be balanced growth in the static model. The extra flux to biomass is shared out between several biomass components rather than all being channelled to trehalose and lipid.

The weighted residual values show that the distribution calculated with the dynamic model is more consistent with all the measurements than the distribution calculated with the static model. The flux distribution patterns at 62 hours are similar. However the weighted residual values show an even greater difference between the two methods with respect to the consistency achieved with the supplied data.

3.3.5.2.1 Summary

This section has shown that in this phase of the fermentation the assumption that biomass composition is constant leads to large errors in the calculation of fluxes to biomass components. However despite this with the exception of the PPP the impact on central metabolic pathways is fairly negligible in this stage of the fermentation. This is probably because the flux distribution through the central pathways is dominated by the OUR and CER values. The fluxes through the central metabolic pathways are so large that even large changes in the flux through peripheral pathways have little effect on the size in the central pathways. The OUR and CER data is measured independently of the biomass composition and so is the same for both calculations. This leads to the reasonably high degree of similarity seen between fluxes in central metabolism using the two methods. Interestingly this is similar to what Daae and Ison (1999) found for *S. lividans*. They found that the central metabolism of *S. lividans* was most sensitive to changes in the OUR. They also found that the PPP was the most sensitive part of

central metabolism to changes in measured fluxes and this too seems to be confirmed by these findings.

3.3.5.3 Comparison of Flux distributions in the Stationary Phase After Glucose Depletion.

In the stationary phase after glucose depletion the metabolic activity drops considerably including the OUR and CER. The use of several carbon sources other than glucose leads to a very different flux distribution through central metabolism than seen before glucose is depleted. The OUR and CER no longer have such a dominating influence on the flux through central metabolic pathways. A comparison has been performed for 90 hours and the distributions are presented as Figure 28 and Figure 29 for the dynamic and static models respectively.

The flux distributions at 90 hours show much greater differences than those seen in earlier phases of the fermentation. There are large differences throughout glycolysis and the PPP. There are also differences in the uptake of α-ketoglutarate, secretion of NH_3 and in fluxes to and from biomass. These differences all stem from the need to use biomass components to satisfy the organism's carbon requirements. When a constant biomass composition is assumed the model is unable to satisfy these requirements.

In both models the PPP shows reversed flux. Negative flux is possible through the non-oxidative branch but is thermodynamically unfavourable in the oxidative branch. Both distributions end up with these negative fluxes for the same reason. When biomass is broken down through the normal catabolic pathways the reduced cofactor produced is NADH. When biomass components are synthesised NADPH is the main reduced cofactor used. If biomass components are broken down in the model using the biosynthetic route then NADPH is generated. This NADPH has to be reconsumed to maintain the steady state. This is done by driving the pentose phosphate pathway backwards. Clearly this is an artificial situation.

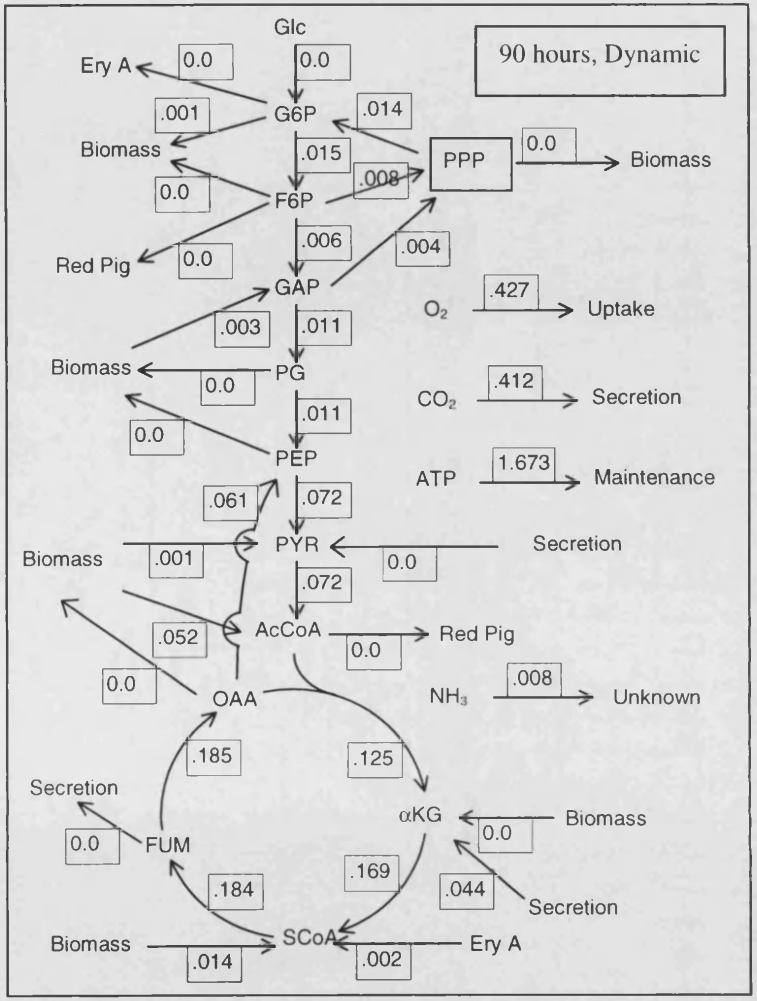


Figure 28: White variant wild type flux distribution at 90 hours based on measured data. Weighted residual = 0.23. Flux to NH_3 undefined. All fluxes are in mmole/gDCW/hour

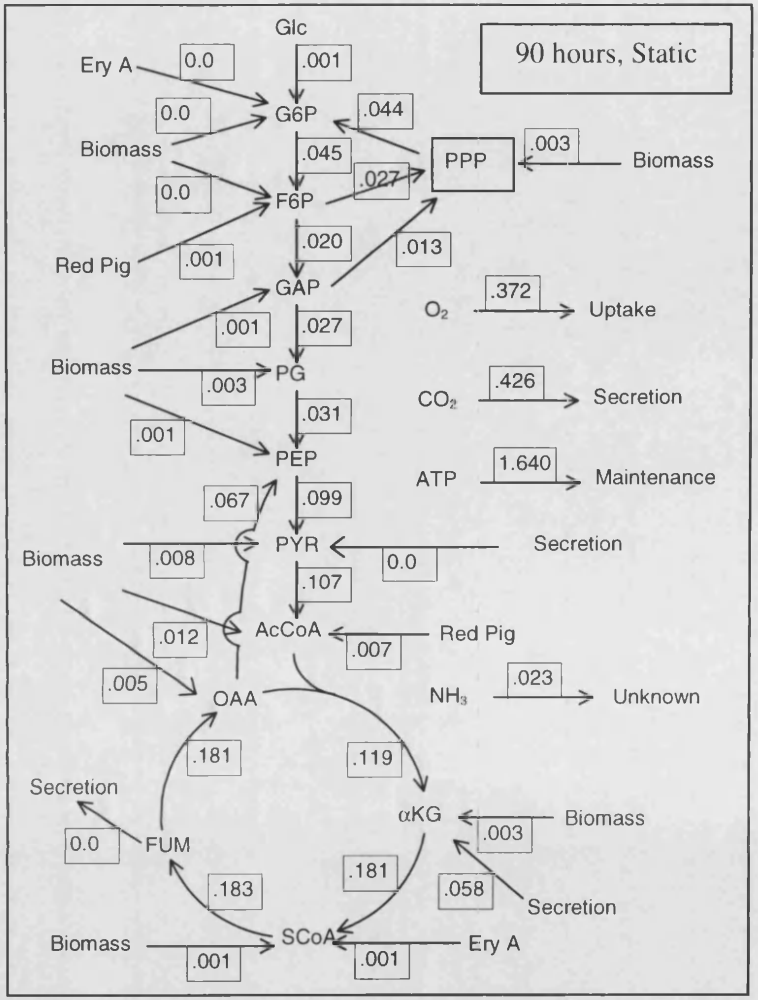


Figure 29: White variant wild type flux distribution at 90 hours based on assumed data. Weighted residual = 3.0. Flux to NH_3 open. All fluxes are in mmole/gDCW/hour

The catabolism of most biomass components follow different pathways than the synthesis, however these pathways are not available to the static model. The static model assumes that biomass composition is constant and as there is no net loss of biomass there is then no need for biomass degredation pathways. In the static model the situation of biomass catabolism is not anticipated and the only pathways available to degrade biomass components are the biosynthesis pathways. This leads to high levels of artificial NADPH production and a large negative flux through the PPP. The flux distribution calculated with the dynamic model also has a small negative flux through the PPP. This is due to the least squares method used to calculate the flux distribution in the network. The model is given the fluxes in these precursor biosynthesis pathways as zero. However as it fits the data it adjusts all the given rates slightly to achieve a consistent set of data. This leads to very small negative fluxes in some of these biosynthetic pathways. In this situation when metabolic activity is so low this can have a detectable impact on the flux distribution through the PPP.

Artificial reversal of the pentose phosphate pathway leads to carbon fixation in the static model. The carbon then passes down through glycolysis and is released in the TCA cycle where NADH is formed rather than NADPH. NB this use of NADH by isocitrate dehydrogenase is a property of the model in the stationary phase not a property of the organism see section 2.10.4. This fixation of carbon is erroneous and has a large impact on the fluxes through the PPP and glycolysis. For example it causes much higher fluxes through glycolysis than seen in the dynamic model.

In the static model there is no facility available to release carbon stored in the biomass. However in order to sustain the measured CER the utilisation of carbon is required. The model supplies this carbon by altering the size of some of the measured rates to achieve consistency. The growth rate is adjusted by the modelling process to be negative in order to supply extra carbon to the network. In a similar manner the uptake of α-ketoglutarate is increased. This increase in the α-ketoglutarate uptake causes increased flux from oxaloacetate to phosphoenolpyruvate. The balanced catabolism of biomass leads to high levels of protein and RNA catabolism compared with the measured values. This leads to a larger flux for NH_3 secretion than seen in the dynamic model.

The large differences in flux distribution is matched by a large difference in the value of the weighted residual calculated. The residual for the static model is much higher than that for the dynamic model. It can be concluded that in this phase use of a fixed biomass composition does not allow a reasonable flux distribution.

3.3.5.4 Consideration of Which Modelling Method is Most Suitable for Studying Erythromycin Production

Observations that the OUR and CER exert a great deal of control over central metabolism but little control over peripheral metabolism raises a question. Which fluxes are important for studying secondary metabolism? If the central pathways are the important ones then it might be less important which method is used. If the peripheral pathways are more important then it will matter which method is used.

Central metabolism is important for erythromycin production because it provides for the supply of all the precursors and cofactors needed for the synthesis of erythromycin. However although central metabolism is fundamental to all cellular functions most of the specifics of erythromycin synthesis occur outside central metabolism. The details of what happens outside central metabolism are of great interest, some of the reasons for this are given below.

1) The pathways of antibiotic synthesis are not central. For example the PKS activity is not tightly tied to the OUR and the CER. It is important to be able to assess flux through these pathways accurately. 2) The pathways supplying precursors to antibiotic synthesis are peripheral. For example the propionyl-CoA and methylmalonyl-CoA precursors can be supplied by the catabolism of valine, isoleucine and threonine or can be supplied by methylmalonyl-CoA mutase from succinyl-CoA. The proportions of these sources used is of interest if high yields are to be attained (see section 4.3.6). Accurate assessment of the relative use of these different potential pathways will be important for studying erythromycin production. 3) Synthesis of erythromycin requires a source of nitrogen. In the stationary phase in nitrogen limited culture this comes from the degradation of biomass components through pathways that are not part of central metabolism. 4) If maximal theoretical yields are to be achieved erythromycin

biosynthesis will have to dominate the flux distribution in the central metabolic pathways rather than the OUR and CER. This might cause the shape and definition of central metabolism to change somewhat. 5) Feeding strategies for increasing production such as supplying propionate or valine in the medium bypass central metabolism altogether but are still of interest for modelling. 6) One important goal to achieve increasing flux to erythromycin will be channelling precursors to erythromycin rather than to competing products. Lipid synthesis competes for propionyl-CoA and catabolic products of valine and isoleucine. Trehalose and arabinogalactan synthesis compete for glucose-1-phosphate. These branch points fall outside central metabolism but they are important for understanding erythromycin production. On the basis of these arguments I maintain that in the study of flux distributions relevant to erythromycin synthesis it is vital to consider both central and peripheral parts of the metabolic network. For this reason it is important to determine the fluxes in all parts of metabolism as accurately as possible.

3.3.5.5 Comparison with the Results of Daae and Ison

Daae and Ison (1999) investigated the impact of a 20% change in the demand for each of the precursors and cofactors involved in biomass synthesis on the flux distribution in a network for *S. lividans*. They found that the response to a 20% change in one of the biomass components in most cases changed the fluxes in the network less than 0.15 percent on average. The largest average change found was 1.14%. They also recorded the largest individual change in the fluxes in the network. In most cases the largest change in the network was less than 1%. The largest change found in all the changes investigated was 7.72%.

The results found in this chapter can be presented in a similar manner. The results are not directly comparable because Daae and Ison were only changing one biomass component at a time. In this work it is the whole biomass composition that is being considered.

Another problem with comparing the results arises from defining the extent of central metabolism. In this work the definition has been restricted to glycolysis, the TCA

cycle, PPP and oxidative phosphorylation. This was done because all other pathways are leading directly to or from biomass components or products. At different times in the fermentation various biomass components are being synthesised or catabolised. In the dynamic model different pathways are used for these different situations which means that some pathways are not used under certain conditions. Central metabolism was defined to cover those pathways that are always used. Daae and Ison used a wider definition of central metabolism which included the pathways for amino acid and nucleotide biosynthesis.

Statistics from Daae and Ison (1999) and from the results found in this chapter are presented in Table 15. For Daae and Ison the results given are the percentage change in fluxes for a 20% change in one biomass component. For experiments performed in this chapter the percentage change between results calculated with the dynamic and static models are given. See section 3.2.1 for details on how these calculations were performed.

Table 15: Comparison of Results With Those of Daae and Ison

	Central Metabolism		Entire Network	
	Average Difference (%)	Maximum Difference (%)	Average Difference (%)	Maximum Difference (%)
Daae and Ison	0.06-1.14	0.06-7.72		
16 hours	9.6	20.8	22.9	737.5
26 hours	1.4	3.2	6.6	100.0
36 hours	1.4	7.3	10.6	128.3
52 hours	6.1	30.2	460.7	4900.0
62 hours	5.0	32.6	567.4	16700.0
90 hours	91.9	226.2	274.6	2100.0

If measuring the biomass composition does improve the accuracy of the flux distribution then the results for the exponential phase should be the most similar to those of Daae and Ison and those for the stationary phase should be significantly worse.

Of the exponential phase results those for 26 and 36 hours are the closest to the results found by Daae and Ison. As noted earlier the results at 16 hours are probably adversely effected by the low biomass concentration at this time leading to inaccuracies in the measurements taken. However although the exponential phase results are the closest to

those found by Daae and Ison they are still somewhat worse. Even when restricted to central metabolism the results for 26 and 36 hours are only comparable with the worst of the results found by Daae and Ison. The worst of their results however were still considered to be acceptable. Daae and Ison considered each biomass precursor individually whilst here all biomass precursors are considered together. This is probably the reason why the differences found for the exponential phase central metabolism are greater than those found by Daae and Ison. The individual differences in precursor demand are probably considerably less than 20% at 26 and 36 hours between the dynamic and static models. However the summed effect of all the differences is greater than a 20% change in any one component.

The results for the entire network at 26 and 36 hours show larger differences. However care has to be taken with interpreting these results. For example at 26 hours the highest change found was 100% however this is based on the change of a flux that was only reported by the model to 1 significant figure. The apparent large change could be the product of rounding. The highest change found which is unlikely to be the product of rounding errors is 27%. Even taking this into account the results outside central metabolism are still considerably worse than those in central metabolism. This is probably because the fluxes tend to be much smaller in peripheral metabolism. This means that small changes in biomass composition have a bigger relative impact on the size of these fluxes. Central metabolic pathways carry such high levels of flux related to respiration and redox balance that small changes in biomass composition have little impact on them.

The results for the stationary phase before glucose depletion show much larger differences than those for the exponential phase. In fact both central metabolism and the whole of metabolism show much greater differences than the results found by Daae and Ison. The reasons why the differences found between the dynamic and the static model are greater in this phase have been outlined earlier. This analysis does however reveal just how big some of the differences are. For 62 hours the biggest difference is for the rate of synthesis of glutamate from α-ketoglutarate which is 167 times higher in the static model than in the dynamic model. This is due to the static model considering that protein is still being synthesised in the stationary phase as biomass accumulates.

The dynamic model on the other hand is using the reaction in the reverse direction to deaminate amino acids as protein is broken down. By this stage however in the fermentation this is a very small flux.

At 90 hours the results show far greater differences. Even the results for central metabolism in the static model on average show nearly 100% difference from the results found with the dynamic model. The assumption of constant biomass composition is completely unsuitable for this time point.

3.3.5.5.1 Summary

A comparison of the results found by the dynamic and static models shows that the static approximation of biomass composition yields results that are far worse than those found with Daae and Ison's approximation. The results found differ most greatly in the stationary phase. The results are not directly comparable as the two experiments were conceptually different. However using their method of comparison it is clear that outside the exponential phase the errors associated with the assumption that biomass composition can be treated as constant grow considerably.

3.4 Conclusions

3.4.1 Capsule

In this chapter *S. erythraea* has been shown to produce a capsule during the stationary phase. The capsule is formed during a period in the fermentation when protein and RNA are being catabolised with no known sink for the nitrogen produced. Based on this and the identification of one of the appropriate genes the capsule is proposed to be composed of polyglutamate.

3.4.2 Biomass Composition

This chapter has provided the most comprehensive biomass composition for *S. erythraea* available to date. It has found interesting patterns in the biomass composition

during the course of a nitrogen limited fermentation. During the exponential phase balanced growth occurs leading to a constant biomass composition. During the stationary phase protein and RNA are broken down whilst lipid and trehalose accumulate and a capsule is synthesised. Once glucose is depleted from the medium the trehalose and lipid are catabolised.

Careful analysis of the biomass composition allowed a number of unforeseen components to be identified. These include trehalose which had not previously been identified in this species. Arabinogalactan was known to be present but was not known to represent a substantial proportion of the biomass. The presence of a capsule in *S. erythraea* was also unknown and its identification was driven by observations of changes in the biomass composition. Treating the biomass of the cell in this way led to a much deeper understanding of the cell and the changes it undergoes between growth and stationary phase.

3.4.3 Impact of Biomass Composition on Flux Balance Analysis

The balanced growth seen in the exponential phase leads to a constant biomass composition. This makes it reasonable to assume a constant biomass composition in this phase. Results obtained using this assumption in the exponential phase are little different from those based on measurements of the biomass composition. This is in good agreement with the findings of Daae and Ison (1999).

By contrast, during the stationary phase before glucose depletion the biomass composition is not constant. Under these conditions assuming a constant biomass composition leads to erroneous calculation of peripheral rates in the metabolic network, especially those closely related to biomass synthesis. The flux distribution in central metabolism is also adversely effected although this is mainly only seen in the PPP. This limitation in the impact of using the static model is due to the domination of the OUR and the CER over the flux distribution between the central pathways.

Under the conditions of carbon shortage that occur after depletion of glucose the biomass composition continues to change as carbon storage compounds are

reconsumed. The lower general rate of metabolism means that these changes in biomass composition have a much bigger impact on the flux distribution than the changes seen in the previous phase. Assumption of a constant biomass composition in this phase leads to large discrepancies in the flux distributions throughout the metabolic network. The static model uses many pathways in thermodynamically unfavourable directions. Using a fixed biomass composition in this phase leads to much larger inconsistencies within the data used to calculate the flux distributions. This leads to very unlikely flux distributions.

The conclusion of Daae and Ison (1999) was that biomass composition does not have a large impact on the fluxes found in their metabolic network for *S. lividans*. Thus the biomass composition of *E. coli* could be used which meant that the biomass composition did not have to be found for *S. lividans*. This finding has been used to justify the use of *E. coli* biomass compositions in a number of metabolic models of *streptomyces* organisms (eg Jonsbu *et al.* 2001 Avignone Rossa *et al.* 2002). Here it has been shown that this simplification cannot be generally applied without care. The biggest changes in demand for biomass precursors come from the peculiar stationary phase metabolism of these organisms themselves. The assumption of a constant biomass composition even if it is accurate for the exponential phase will not give accurate results for the stationary phase of the fermentation. Recently Roubos (2002) has also challenged the conclusions of Daae and Ison (1999) on the basis they considered only one monomer at a time and do not take into account the large biomass variations found in *streptomyces* species. He did not however go on to demonstrate the implications as has been done here.

In this chapter it has been shown that when studying the stationary phase of *S. erythraea* using flux balance analysis the changing biomass composition needs to be taken into account. The methods for measuring the components have been put in place and model and methods for calculating the flux distributions have been established. The methods developed will be used in the next two chapters to investigate the changes occurring between growth phase and stationary phase (chapter 4) and to investigate the differences between an industrial strain and its wild type parent (chapter 5). The ability to accurately determine flux distributions is essential to the work of these chapters.

4 COMPARISON OF METABOLISM IN GROWTH PHASE AND STATIONARY PHASE

4.1 Introduction

In chapter 3 methods were developed to enable flux balance analysis to be used to study the stationary phase of *S. erythraea*. This allows the behaviour of *S. erythraea* during the stationary phase to be investigated using flux balance analysis for the first time. In this chapter the method developed in chapter 3 will be used to study the metabolism of *S. erythraea* during all stages of the fermentation. The metabolic properties of the different stages will be compared and discussed.

One of the interesting features of organisms like *S. erythraea* is the radically different modes of metabolic operation they can display in different phases of a fermentation and the speed with which they can move between these phases. Under the batch culture conditions used in this work three distinct phases can be seen. A phase when nutrients are unlimited which is dominated by growth. A phase when nitrogen is limiting but carbon is abundant which is characterised by the production of non-growth associated products such as organic acids and erythromycin. Finally there is a phase when both primary carbon and nitrogen sources are used up. During this phase metabolism proceeds at a low level supported by internal stores of carbon and organic acids produced in the second phase.

If the metabolism of *S. erythraea* is to be manipulated it is important to understand what factors are controlling the metabolism. The growth phase in *S. erythraea* is similar to those seen in other organisms. It is driven by the need to quickly produce precursors of biomass for growth. However despite a lot of speculation over the years little is understood about what drives the peculiar metabolism seen in the stationary phase of *S. erythraea* and similar organisms. This second phase is commercially the most interesting as it is the phase in which secondary metabolites tend to be produced. A better understanding of what is happening in this phase might enable strategies to be devised for the overproduction of secondary metabolites such as erythromycin.

When trying to understand the behaviour of a complex system it helps to be able to quantify its behaviour in order to compare less well understood systems with more well understood systems. By performing flux balance analysis it is possible to quantify the behaviour of metabolism during the three phases of growth seen in batch culture. It is then possible to compare the flux distributions seen under the different conditions and to start to speculate on the causes of the differences. Once new ideas about the causes have been formed strategies can be devised to test these ideas.

Batch culture was chosen over continuous culture for these studies because batch and fed batch fermentation are the forms of cultivation most commonly used in industry. It is also more or less impossible to simulate the different phases of a fermentation process in continuous culture because the culture is unstable at the maximum growth rate (μ_{max}) which represents the exponential growth phase and is unable to operate at a growth rate of 0 representing the stationary phase.

In this chapter flux distributions from the different phases of a fermentation will be compared to reveal any significant differences. From these differences ideas will be formulated regarding the factors controlling flux distributions in the stationary phase. Comparison will be made with theoretical flux distributions which give the maximum possible yield of erythromycin on glucose. Important differences will be identified and the implications considered.

4.2 Materials and Methods

All materials and methods used in this section are as stated in the main materials and methods section. Fermentation was performed in the New Brunswick bioreactors described in section 2.3. The strain used was *S. erythraea* wild type. The fermentation was performed in duplicate and very good reproducability was seen. Samples were taken regularly throughout the fermentation and analysed for: DCW, glucose, nitrate, α-ketoglutarate, pyruvate, fumarate, erythromycin, red pigment, total organic carbon, DNA, RNA, protein (intra and extracellular), trehalose, arabinogalactan and lipid.

4.2.1.1 Identification of fumarate

Fumarase and sodium fumarate were obtained from Sigma. 100 mM tris pH 7.6 was prepared in ultra-pure water. A standard solution of 0.5 g/l fumarate was prepared in ultra pure water. Fumarase was made up to 2 units/µl in 50% glycerol. Sample (0.25ml) or standard (0.25 ml tris + 0.06 ml fumarate) were incubated at 25°C for 30 minutes with 0.001 ml of the fumarase solution. Controls used ultra pure water instead of fumarase.

After incubation fumarate was determined without further treatment by HPLC using the organic acids protocol given in section 2.6.15.2. Reduction in the level of fumarate in the experimental sample compared to the control was taken as an indication that the peak was fumarate.

4.2.1.2 Elementary modes analysis

Elementary mode analysis was performed on the metabolic network described in section 2.10. The modes giving the highest yield of erythromycin on glucose and RNA and on glucose and protein were found and compared with the flux distributions found in the wild type organism. Elementary modes analysis was performed in Matlab version 6.0.0.88 using Flux Analyzer version 4.1 (Klamt *et al.* 2003). In order to reduce computational time only subsets of the metabolic network were used. Pathways for RNA, DNA, Lipid, arabinogalactan, trehalose and most amino acids were not used as it was considered unlikely that the highest yield of erythromycin on glucose and RNA or protein would involve the synthesis of biomass. Pathways for the synthesis of valine, isoleucine and threonine where however used as these amino acids can be catabolised to produce precursors of the polyketide moiety of erythromycin. Pathways for production and secretion of organic acids were not used because it was thought unlikely that secretion of carbon would be compatible with high yields of erythromycin. Gluconeogenesis was thought unlikely to play a role so phosphoenolpyruvate carboxykinase was not included. Malic enzyme and the glyoxylate bypass were not included in order to reduce the computational time. Given the demand for NADPH for erythromycin synthesis malic enzyme which consumes NADPH when supplying

anaplerotic requirements was thought unlikely to have a role in the highest yielding modes.

Methylation reactions have an important role in erythromycin biosynthesis. The exact mechanism by which tetrahydrofolate (THF) is charged with a one carbon unit is unknown in *S. erythraea.* A simplistic assumption concerning this reaction (reaction 62) was used by Daae and Ison (1999) and Ushio (2003). This simplification was maintained here and was satisfactory for use in flux balance analysis where fluxes through the reaction were always low compared with other fluxes in the network. When the network was used for elementary modes analysis however this simplification was found to be unsatisfactory as the flux through this reaction in the modes with the highest erythromycin production were quite high. In this situation the reaction had a large impact on the flux in the rest of the network. For the elementary modes analysis this pathway was change to:

$$CO_2 + NADPH + THF + ATP \rightarrow NADP^+ + FTHF + ADP$$

This assumes that CO_2 is reduced to formate via formate dehydrogenase, this formate is then used to generate N^{10}-formyl-THF from THF via N^{10}-formyl-THF synthase.

4.3 Results and Discussion

4.3.1 Fermentation Profiles

The data collected from two fermentations was used to calculate uptake rates for use in the model. The fermentations showed a high degree of reproducibility and only one set of results is presented here. The data is presented as fermentation profiles in Figure 30 to Figure 32.

Figure 30 shows the OUR, CER, glucose, nitrate and α-ketoglutarate profiles. The culture is nitrate limited and the nitrate profiles show that the nitrate runs out at about 28 hours. This marks the first major switch in metabolism. Growth halts, the CER drops by around a third and organic acid production starts. Following this the OUR and

the CER continue to gently fall before rising again to a new peak at about 35 hours from which point they gently fall until the glucose runs out.

There is an unusual feature in the glucose profile. It shows an early rise before taking on the expected profile. The reason for this is unknown but it is seen consistently with the HPLC method and makes it impossible to use the first 15-20 hours of glucose data for modelling. However as the model is overdetermined the glucose uptake can be found from the other rates at these early time points. The uptake rate of glucose increases until the nitrate runs out. It then holds reasonably constant until about 45 hours. The uptake rate then slows slightly as the glucose runs out.

The depletion of the glucose causes the second big shift in the profiles. The OUR and the CER drop by about two-thirds to a very low level suggesting basal metabolism. At this point the α-ketoglutarate starts being taken up.

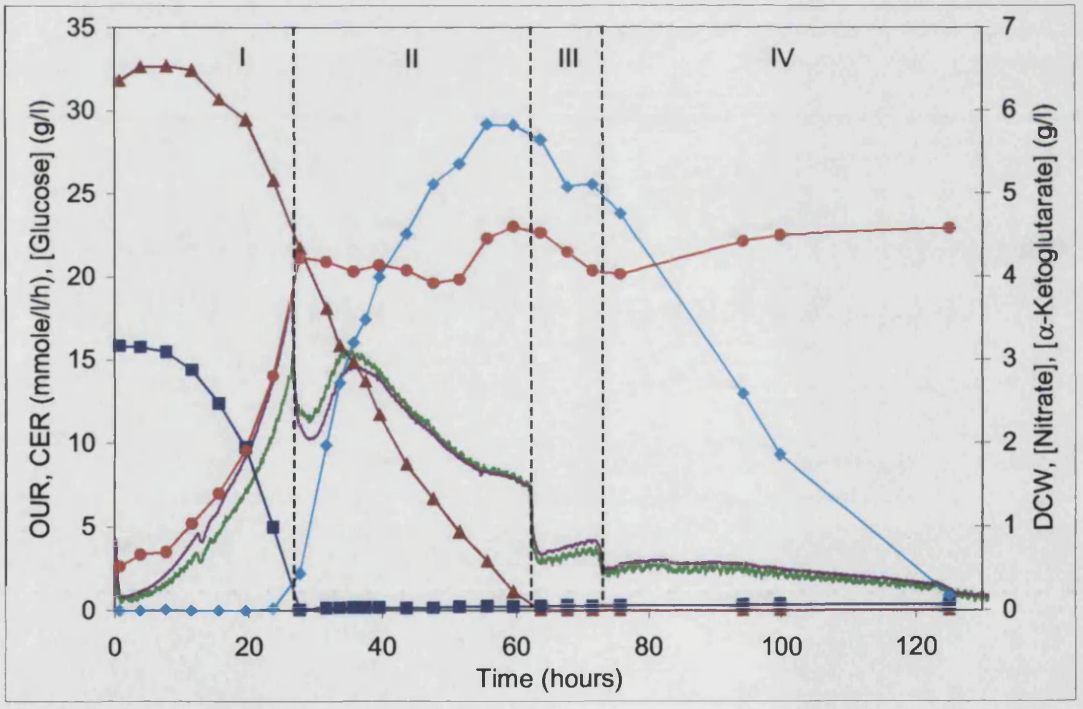


Figure 30: White variant wt fermentation 4, profile 1. OUR (mmole/l/h) — , CER (mmole/l/h) — , Glucose (g/l) ▲ , Nitrate (g/l) ■ , α-Ketoglutarate (mono sodium salt) (g/l) ◆ , Dry cell weight (g/l) ● . Four phases of the fermentation are shown by vertical dashed lines.

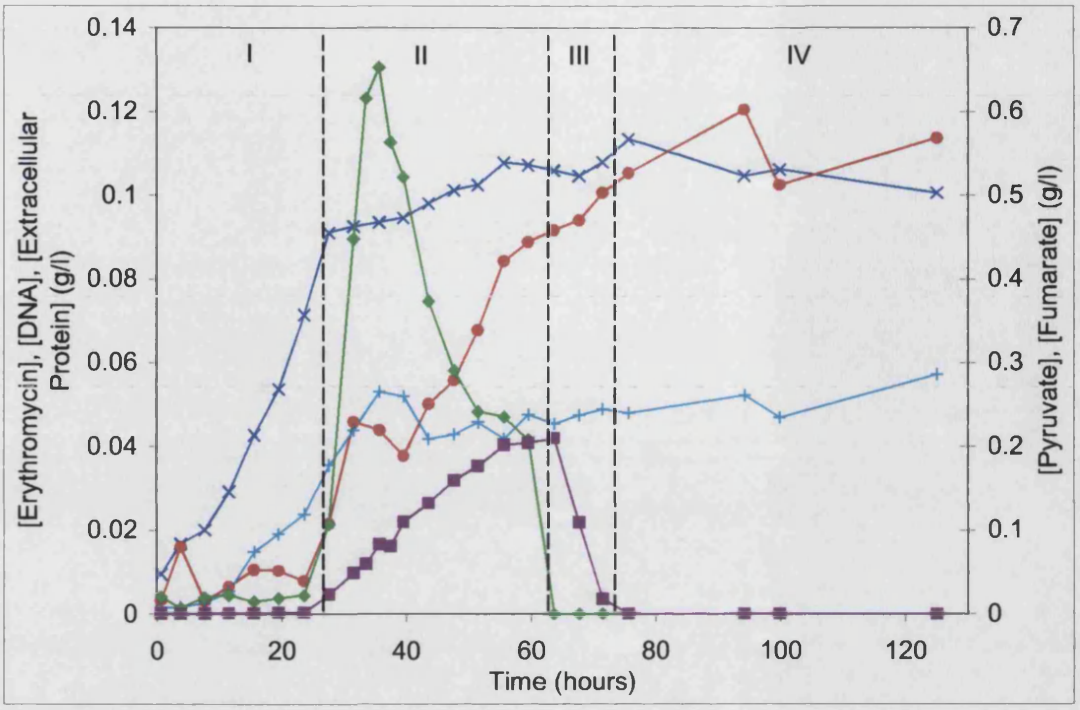


Figure 31: White variant wt fermentation 4, profile 2. Fumarate disodium salt (g/l) ■ , Extra cellular protein (g/l) × , Erythromycin (g/l) ● , DNA (g/l) + , Pyruvate (g/l) ◆. Four phases of the fermentation are shown by vertical dashed lines.

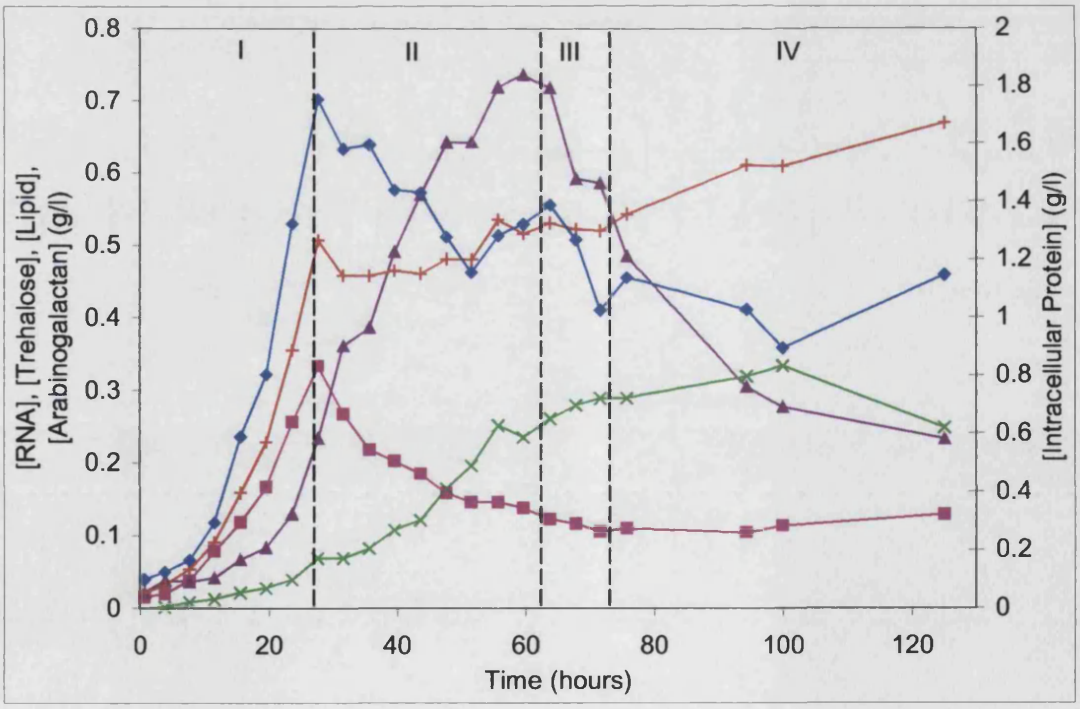


Figure 32: White variant wt fermentation 4, profile 3. RNA (g/l) ■, Lipid (g/l) ×, Trehalose (g/l) ▲, Arabinogalactan (g/l) +, Intracellular Protein (g/l) ◆. Four phases of the fermentation are shown by vertical dashed lines.

During the second phase there does not seem to be any growth, the dry cell weight remains reasonably constant. This is rather different compared with some of the earlier fermentations where there was some accumulation of biomass in this phase. Although there is no apparent increase in DCW there was considerable wall growth in this phase and this has a proportionally larger impact in these vessels which are smaller volume than those used in the earlier experiments. This may explain the difference seen.

Figure 31 contains the pyruvate, fumarate, extracellular protein, erythromycin and DNA profiles. The pyruvate and fumarate profiles show some similarities to the α-ketoglutarate profile. Both start to be produced at the same time, when the nitrate runs out, and both are taken back up by the cell. The levels produced are however much lower. The pyruvate starts being taken up before the glucose runs out and the process is abruptly brought to completion when the glucose runs out. The fumarate is rapidly consumed over 8 hours when the glucose runs out. When the fumarate is depleted the OUR and CER fall further. However it is not clear that this can be directly attributed to the utilisation of the fumarate because when uptake of fumarate stops uptake of α-ketoglutarate increases by about the same amount. This phenomenon with the OUR and CER is routinely seen in fermentations under these conditions with the white variant wild type. Extracellular protein builds up in the fermenter until the nitrate runs out, from then on there is little increase. This is not surprising as protein contains nitrogen. Erythromycin is produced when the nitrate runs out in a classic non-growth associated profile. Once glucose runs out the production of erythromycin more or less stops.

The DNA profile which is also included on this figure shows much the same shape as the growth curve however it carries on increasing for a short time after the DCW has stopped increasing.

Figure 32 contains the profiles for the biomass components except the DNA. The protein and RNA increase in proportion to the biomass during the growth phase but decrease rapidly once the nitrogen source is exhausted. Trehalose is present throughout the fermentation but accumulates particularly during the second phase. Once glucose runs out the trehalose starts to be reconsumed. Lipid is also accumulated in the second

phase, there is however less evidence of its reconsumption before the end of the fermentation for further discussion on the shape of the biomass profiles see chapter 3.

The fermentation profiles show all the hallmarks of a typical *S. erythraea* fermentation on this media. The data yielded is good enough to derive uptake rates and production rates with a reasonable measure of confidence.

4.3.2 Elemental balances

Flux balance analysis is a form of mass balancing. For this reason it is important to account for as much of the mass as possible in the fermentations. By performing elemental balances it is possible to check whether the measurement of uptake rates and production rates are in close enough agreement for flux balance analysis to be performed. To this end carbon and nitrogen balances were performed on the fermentation.

In the carbon balance the total carbon measured in all the components in the broth is compared with the total carbon in the broth as measured by total carbon analysis of the broth. The summation of the carbon in the fermenter, the carbon released as CO_2 and the carbon removed in samples is presented in Figure 33. The carbon balance closes reasonably well in the earlier stages of the fermentation. However is becomes worse as the fermentation progresses both in terms of the components measured in the broth and the carbon balance over the whole fermenter. Some of the discrepancy between the measured and actual values can be attributed to the biomass measurements because not all components are measured. However values are assumed for these missing components in the model and so this should not effect the calculations. If a capsular material is being formed in the stationary phase this may also account for some of the discrepancy. Further discrepancy is caused by the addition of acid and base to maintain the pH of the broth. This dilutes the broth when compared with the first sample giving the impression that less carbon has been accounted for than actually has. This has an adverse effect on the carbon balance but does not effect the calculation of uptake rates used for modelling.

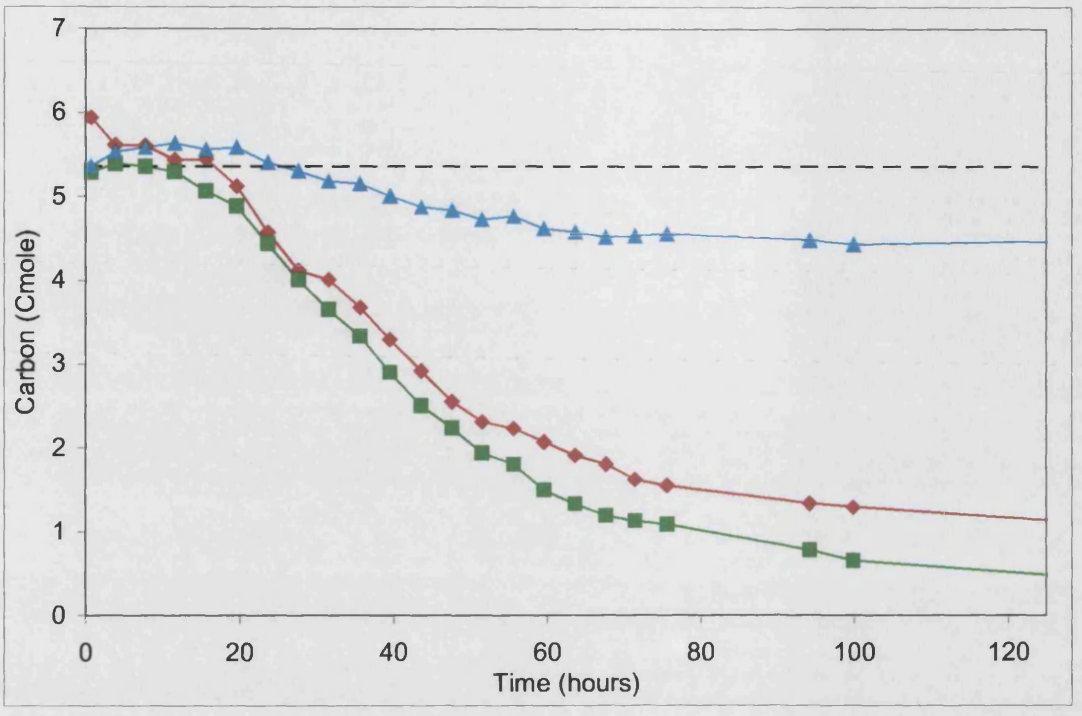


Figure 33: Carbon balance for white variant wild type fermentation 4. Total organic carbon in broth ◆, summed carbon in components measured in the broth ■, sum of carbon in broth, samples and off gas ▲, initial carbon in reactor -------.

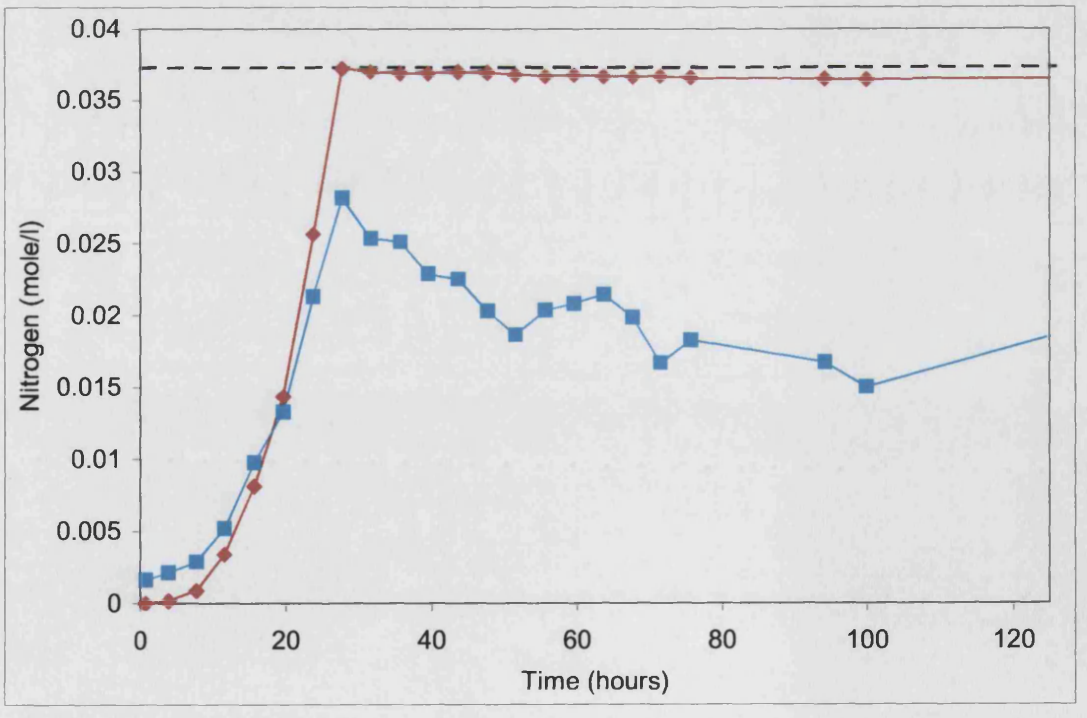


Figure 34: Nitrogen balance for white variant wild type fermentation 4. Nitrate consumed ◆, summed nitrogen in components measured in the broth ■. Initial nitrogen in reactor ------.

In the nitrogen balance the amount of nitrate consumed is compared with the total amount of nitrogen measured in biomass and products. The total nitrogen in measured biomass components and products over the course of the fermentation has been plotted in Figure 34. The consumption of nitrate is plotted along side it for comparison. The nitrogen balance does not close tightly especially in the stationary phase when there is a large discrepancy. This has effected how modelling has been performed. The amount of nitrogen measured in the biomass falls during the stationary phase. In the stationary phase an escape route was provided in the model for this nitrogen as without this there was no satisfactory way for the model to balance the nitrogen in the stationary phase. This was provided in the form of an imaginary flux to extracellular ammonia which the model calculated based on the data. Ammonia was chosen as it was considered that wherever the nitrogen was going it was probably going to be at this level of reduction therefore choosing ammonia should have the minimal impact on the redox balance.

4.3.3 Identification of a New Product

The carbon balance was observed to become progressively worse in the stationary phase (see section 4.3.2). In order to find the destination of the missing carbon an unidentified peak on the organic acids HPLC was identified. Under the conditions used the peak elutes at around 13.9 minutes. Analysis of the peak for total organic carbon revealed that its maximum peak area represented slightly less carbon than the pyruvate peak.

Samples were spiked with a number of organic acids and the peak was found to coelute with fumarate. In order to confirm this finding fumarase enzyme was obtained. Enzymes are very specific regarding the identify of the substrate accepted. Showing that this compound acts as a substrate for fumarase gives good confidence that it is fumarate. Fumarase catalyses the conversion of fumarate to malate in the citric acid cycle. Under physiological conditions the equilibrium position is around 1:4.4 fumarate to malate respectively (Rose *et al.* 1992). This means that an incomplete conversion of fumarate is expected.

A 0.5 g/l solution of fumarate was split in two. One batch was incubated with fumarase the other without. The resulting HPLC chromatograms for these batches have been

superimposed and are presented in Figure 35. When fumarase is incubated with the fumarate solution the fumarate is depleted. The malate produced is not detected however the depletion of the fumarate shows the enzyme's activity.

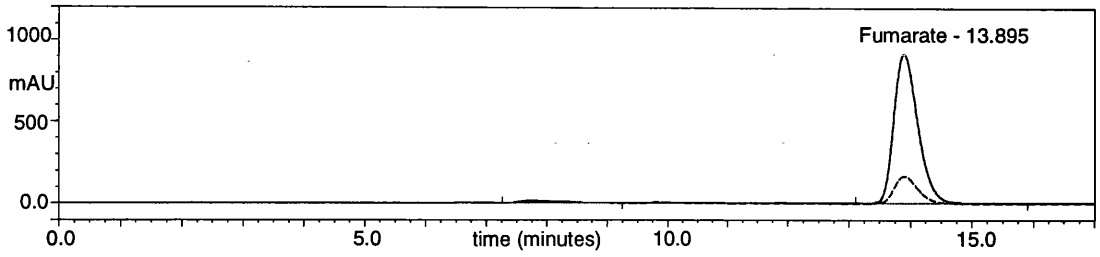


Figure 35: Superimposed chromatograms for fumarate solutions incubated with and without fumarase. Without fumarase——. With fumarase------.

A sample of fermentation broth was split into two batches. One was incubate in the presence of fumarase and one without fumarase. The samples were then analysed for fumarate by HPLC and the resulting chromatograms have been superimposed and are presented in Figure 36. In the presence of fumarase there is a significant reduction in the proposed fumarate peak compared with the control. This is very strong evidence that the peak is fumarate. In the standard solution the conversion approaches the expected ratio of 4.4 malate : 1 fumarate. In the filtered broth the conversion ratio is lower. The pH of the broth is buffered at 7.0 whereas the pH of the standard was buffered at 7.5. The sample also contained higher levels of salts than the standard. The change in the equilibrium position of the reaction is probably a result of these differences.

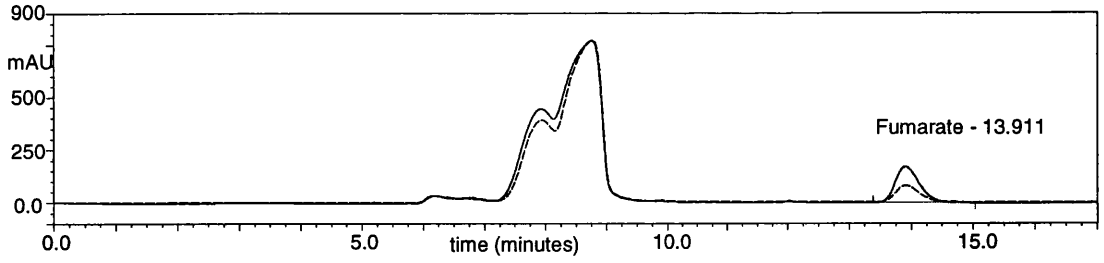


Figure 36: Superimposed chromatograms for filtered fermentation broth incubated with and without fumarase. Without fumarase ——. With fumarase------.

Fumarate is produced by *S. erythraea* during the stationary phase in a manner similar to the other organic acids produced i.e. α-ketoglutarate and pyruvate. Production starts as soon as nitrate is depleted, however when glucose is depleted the fumarate is

reconsumed (see Figure 31). The production of fumarate in a bacterial fermentation is a little unusual. It is not one of the classic growth associated organic acid products. However organic acid production is not growth associated in *S. erythraea.* α-Ketoglutarate and pyruvate are usually associated with overflow metabolism rather than directly linked to growth. Production of fumarate is known in a number of fungi and high yields can be obtained with rhizopus species (Cao *et al.* 1996). To the best of my knowledge this is the first report of the production of fumarate in *S. erythraea* or in fact any actinomycete. Secretion of a number of TCA cycle intermediates and other organic acids has been reported in *streptomyces* species. Kannan and Rehscek (1970) reported that when high levels of glucose were fed to starved cells of *Streptomyces antibioticus* succinic, citric, pyruvic lactic and α-ketoglutaric acids were secreted. This is however somewhat different to the situation reported here. Some speculation on the cause of fumarate secretion will be given in chapter 6.

The levels of fumarate found were quite low in the context of other substrates and products of the fermentation. The identification of fumarate did not lead to large differences in the flux distributions found in the metabolic network. The fumarate data has been included in the carbon mass balance shown previously in section 4.3.2.

4.3.4 Comparison of Flux Distributions in Growth Phase and Stationary Phase.

The fermentation data was used to perform flux balance analysis at various time points through the fermentation. Both fermentations gave similar flux distributions throughout. Six flux distributions representative of four distinct stages in the fermentation are presented as Figure 37 to Figure 42. The figures are presented with the value of the residual (h) which is a measure of how consistent the calculated flux distribution is with the measured data, the smaller the value the more consistent the flux distribution. If any of the measured rates were neglected in the modelling this is also recorded in the figure legend. All flux distributions are taken from one fermentation. The stages chosen are: 1) The growth phase before nitrate depletion. For this phase the flux distribution at 25 hours is shown. 2) The stationary phase before glucose depletion. This phase is the phase in which most antibiotic production occurs. During

it considerable changes in the measured profiles are seen. For this reason flux distributions for three time points in this phase are included, 40 hours, 50 hours and 60 hours. 3) The stationary phase after glucose depletion but before fumarate depletion. For this phase the flux distribution at 67 hours is shown. 4) The stationary phase after glucose and fumarate depletion. For this phase the flux distribution at 90 hours is shown. The period immediately after the nitrogen source runs out showed very interesting behaviour, however it was so dynamic that uptake rates could not be calculated in this period for some of the profiles. For this reason no flux distribution has been calculated for this period. Attempts made for this transitional period however suggest that the distribution is reasonably similar to that seen at 40 hours.

4.3.4.1 Analysis of Reduced Cofactor Balances During Growth and Stationary Phase

The central metabolic pathways of *S. erythraea* consist of: glycolysis, the TCA cycle and the pentose phosphate pathway (PPP) (see Figure 43). The flux distribution in these pathways is very strongly linked to the OUR and the CER which in turn are strongly linked to how the cell is growing. When the cell is rapidly growing on glucose it requires reducing power for two main purposes: energy and growth. Different reduced cofactors are used for each purpose. The majority of NADH and $FADH_2$ are used to produce ATP to supply the cells energy requirements. NADPH is generally used in the reduction of the carbon source to the same level of reduction as the precursors of biomass formation. When the cell is not growing however there is little demand for NADPH as no new biomass precursors are required. The cell however still requires energy for maintenance purposes so there is still a demand for NADH. In this section the changes in reduced cofactor demand during the course of the fermentation will be investigated.

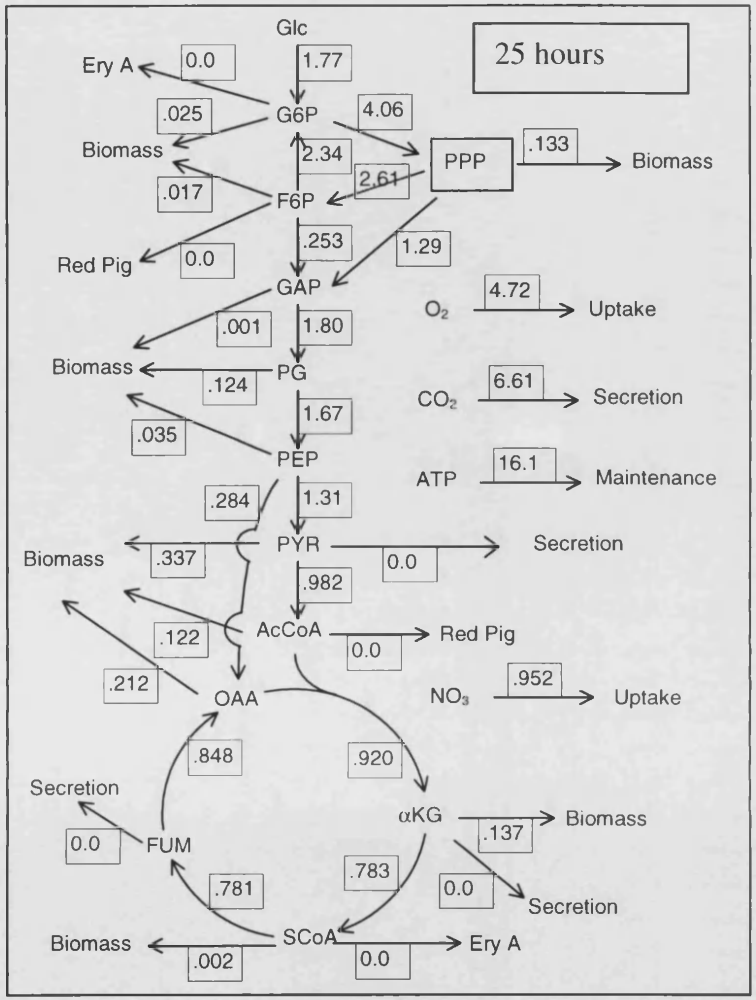


Figure 37: White variant wild type fermentation 4 flux distribution at 25 hours (h = 2.83) CO_2 neglected. All fluxes are in mmole/gDCW/hour

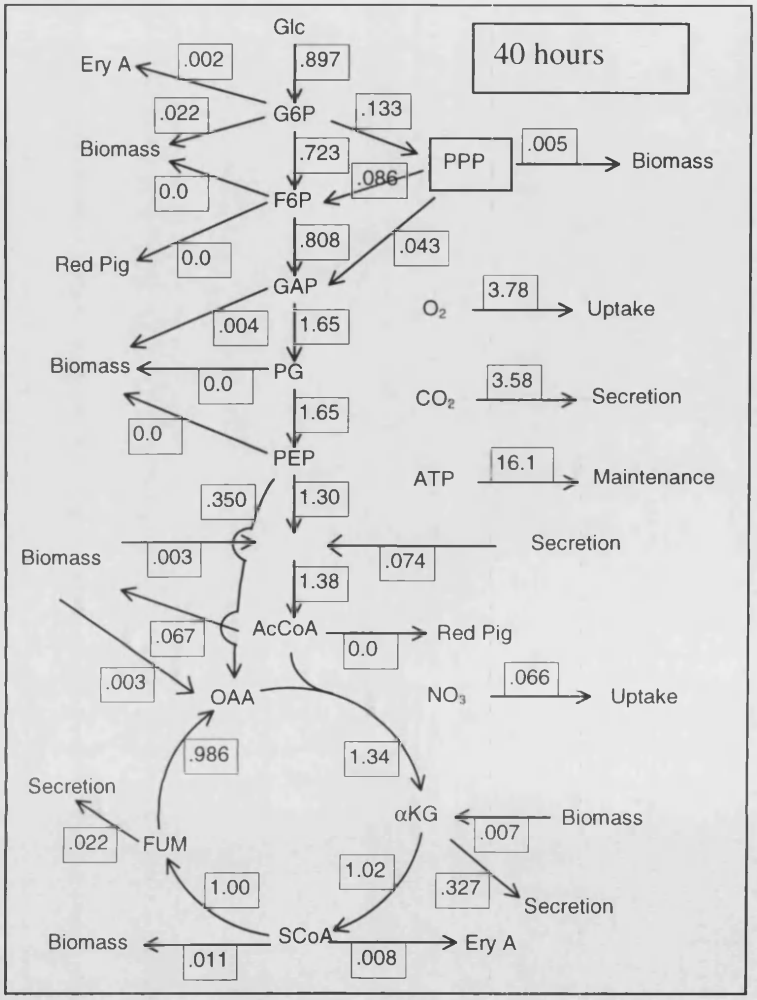


Figure 38: White variant wild type fermentation 4 flux distribution at 40 hours (h = 2.46) NH_3 neglected. All fluxes are in mmole/gDCW/hour

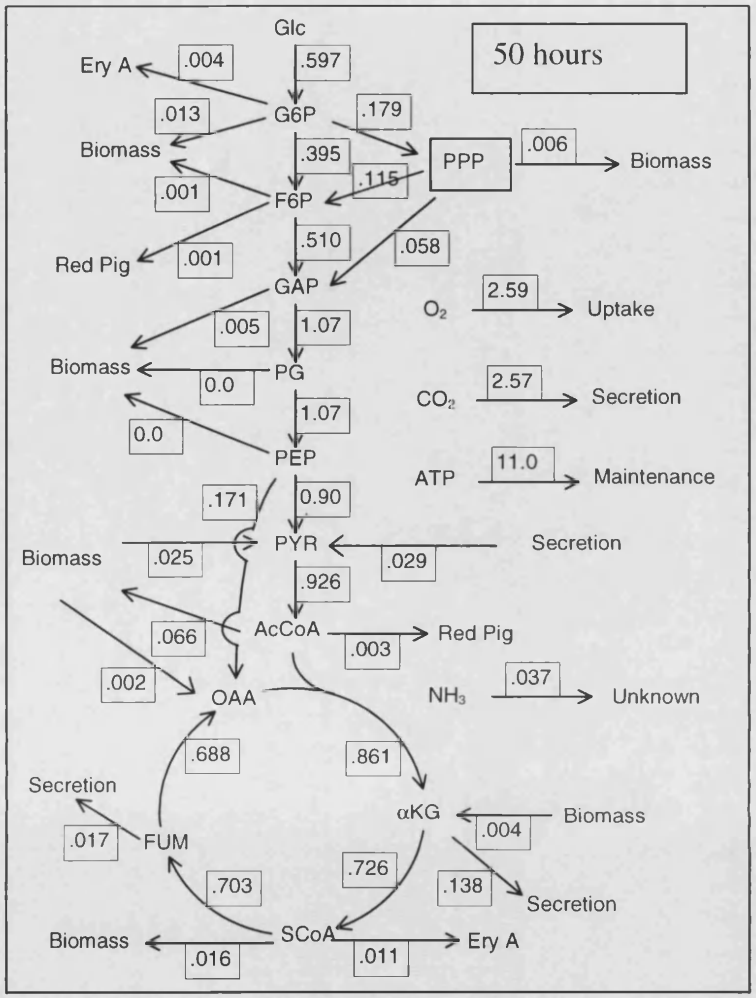


Figure 39: White variant wild type fermentation 4 flux distribution at 50 hours (h = 4.505) NH_3 neglected. All fluxes are in mmole/gDCW/hour

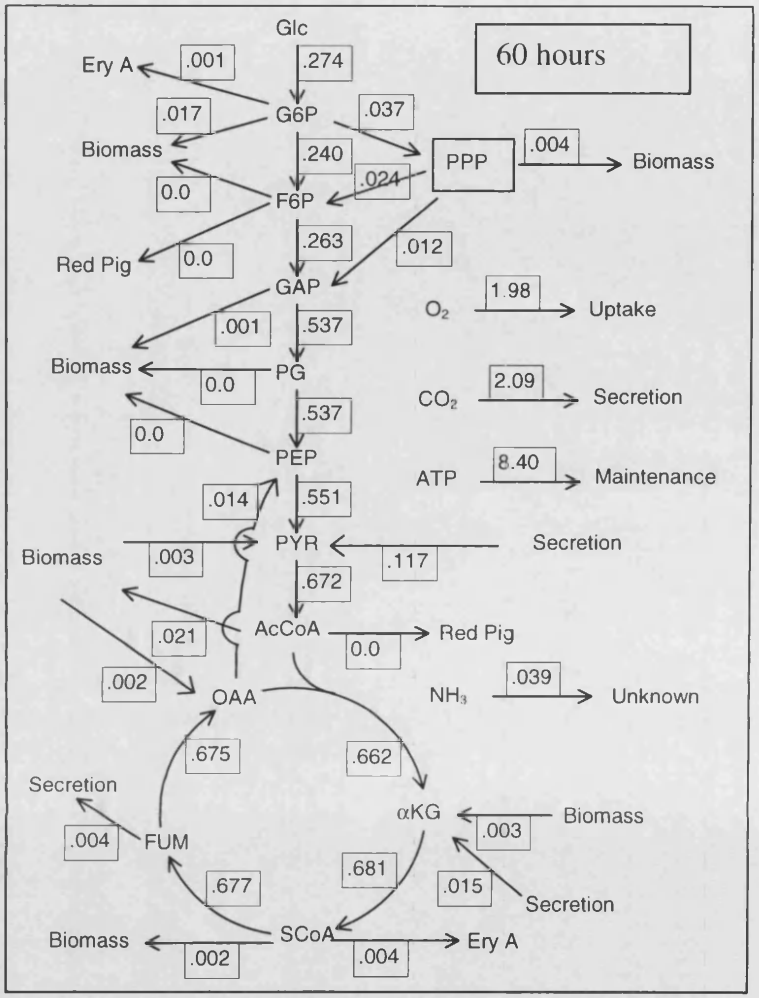


Figure 40: White variant wild type fermentation 4 flux distribution at 60 hours (h = 0.242) NH_3 neglected. All fluxes are in mmole/gDCW/hour

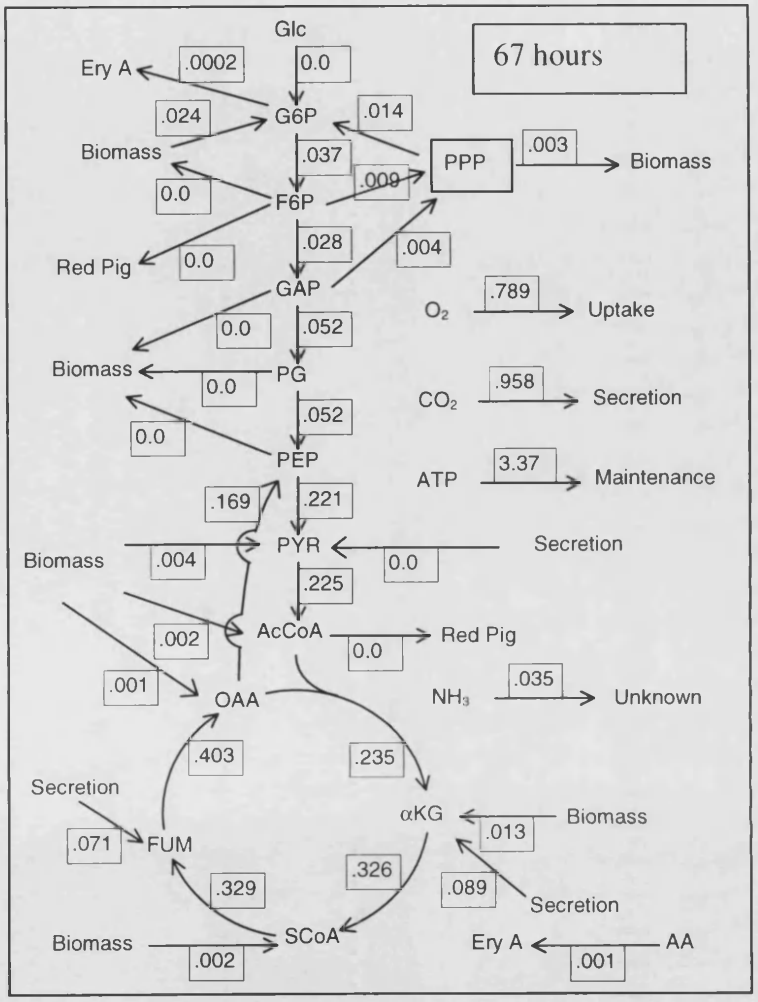


Figure 41: White variant wild type fermentation 4 flux distribution at 67 hours (h = 0.427) NH_3 neglected. All fluxes are in mmole/gDCW/hour

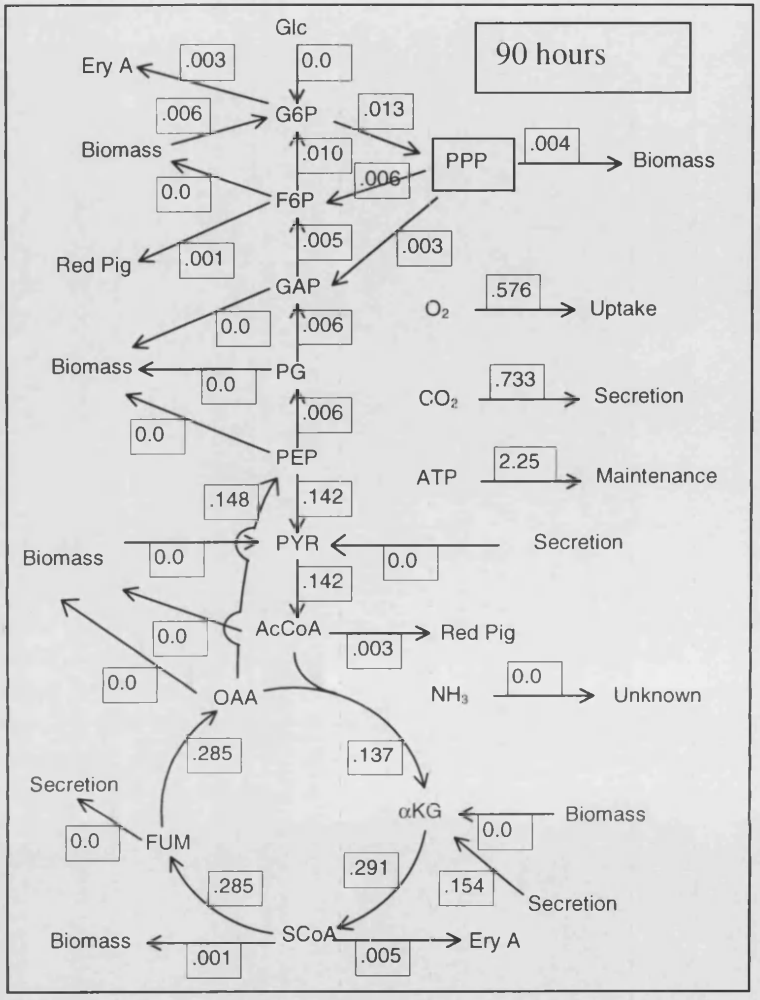


Figure 42: White variant wild type fermentation 4 flux distribution at 90 hours (h = 0.427) NH_3 neglected. All fluxes are in mmole/gDCW/hour

In order to regenerate these cofactors their oxidised forms have to be reduced again. The dehydrogenase reactions which perform this whilst reducing the cofactor concomitantly oxidise a group on another metabolite. A number of these dehydrogenase enzymes oxidise acid groups producing CO_2. Although there are other CO_2 producing reactions in the cell the large demand for reduced cofactors means that the dehydrogenase reactions far outweigh the others. This means that the CER is a reflection of the dehydrogenase activity in the cell and although this corispondence is not complete it serves as a useful guide. The phosphorylation of ADP to ATP by the oxidation of NADH involves the reduction of O_2 to H_2O. If all the reduced cofactors produced by catabolising glucose are used to phosphorylate ADP then the ratio of CO_2 produced to O_2 consumed is 1:1 (see section 6.3.3.2). If NADPH was the only reduced cofactor produced in the catabolism of glucose the ratio would be 1:0. The ratio of the CER/OUR is known as the RQ.

It is possible then from the OUR and CER data to work out the amount of NADH and NADPH being produced in the cell. In *S. erythraea* the PPP produces a ratio of 1 NADH: 1 NADPH. The use of glycolysis and the TCA cycle produces 1 NADPH for every 4 NADH and 1 FADH2 reduced. Combining these ratios with the absolute values of the OUR and CER the fluxes through the central metabolic pathways can be determined. This is one of the constraints used in the model to calculate the flux distribution in the metabolic network.

4.3.4.1.1 Growth Phase

Throughout the growth phase the RQ is greater than 1 because reduced cofactors are being used to produce biomass. This leads to large fluxes though the pentose phosphate pathway (see Figure 37). This is particularly the case in *S. erythraea* as it only produces 1 NADPH per CO_2 released in this pathway unlike many other organisms which produce two (see Figure 43). The demand for NADPH is so high that the pentose phosphate pathway forms a cycle with on average more than 2 carbons per glucose taken up being removed as CO_2 in the PPP. High fluxes through the pentose phosphate pathway have been reported before in related organisms by direct measurement methods. Dekleva and Strohl (1988a) found that the PPP was the main route for carbon catabolism in the growth phase of *Streptomyces C5*. This was determined by

radiolabelling techniques which are unaffected by the presence of transhydrogenase enzymes. Transhydrogenase activity will be discussed later because it can cause problems for the calculation of these fluxes. The flux through glycolysis and the TCA cycle is quite low in comparison with the flux seen through the pentose phosphate pathway reflecting the dual role of the PPP in providing both NADPH and NADH in *S. erythraea*.

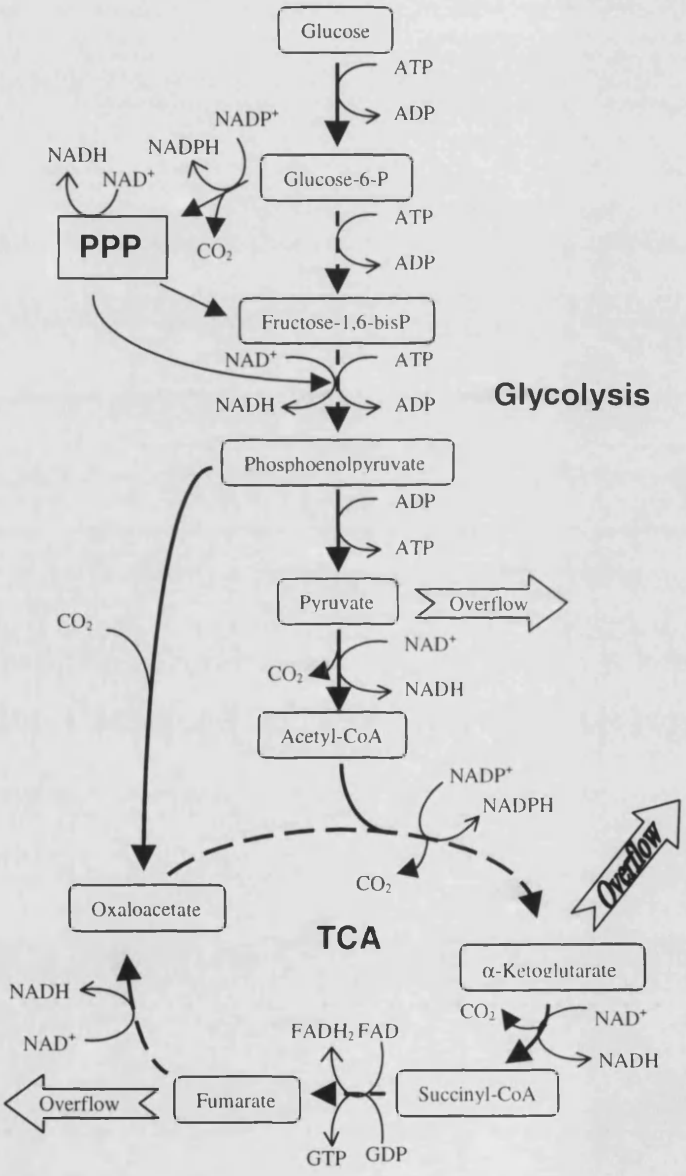


Figure 43: The central metabolic pathways of *S. erythraea*.

4.3.4.1.2 Difficulties with NADPH Balances in the Stationary Phase

When the nitrate runs out the cells enter the stationary phase. There is little growth in this phase and so little demand for NADPH. The demand is so low that it causes difficulties with the modelling. The TCA cycle and the PPP produce different ratios of NADH and NADPH, but both pathways produce NADPH. The flux through these pathways is found by a linear combination of the two that produces the required amounts of NADPH and NADH. This means that there is a minimum ratio of NADPH to NADH that can be produced by these pathways. This ratio is 1:5 for the TCA cycle (counting $FADH_2$ as NADH). If the demand for NADPH falls below this value there is no legitimate combination of these pathways that will achieve the required ratio. The only way it can be achieved is by driving the TCA cycle and the PPP in opposition. NADPH and NADH are produced with a ratio of 1:5 in the TCA cycle and are reoxidised with a ratio of 1:1 by driving the PPP backwards. It is however thermodynamically impossible for glucose-6-phosphate dehydrogenase to work in this direction. Therefore *S. erythraea* must have a different mechanism for avoiding the accumulation of NADPH in the stationary phase.

The most obvious solution is the use of a transhydrogenase enzyme to inter-convert NADPH and NADH. Unfortunately it is not possible to perform flux balance analysis with a transhydrogenase in the metabolic network. Inclusion of a transhydrogenase yields an infinite number of possible flux distributions that satisfy the demands for NADH and NADPH. This is the case because any ratio of PPP and the TCA cycle flux can be made to fit by converting the excess cofactor to the lacking cofactor.

Transhydrogenase activity has not been found in *S. erythraea*. Roszkowski *et al.* (1971) were unable to detect transhydrogenase activity at any fermentation stage. However something similar may still be happening in *S. erythraea*. One possibility is a cycle in which NADPH is consumed and NADH is produced. Roszkowski *et al.* (1971) ruled out one such cycle involving malic enzyme and pyruvate carboxylase by showing that pyruvate carboxylase is not active in *S. erythraea*. However another cycle involving phosphoenolpyruvate carboxykinase is possible. In this cycle pyruvate is carboxylated to malate by malic enzyme and NADPH is converted to $NADP^+$ in the process. Malate is then oxidised to oxaloacetate by malate dehydrogenase reducing NAD^+ to NADH. The oxaloacetate is decarboxylated to phosphoenolpyruvate by phosphoenolpyruvate carboxykinase. The phosphoenolpyruvate is then dephosporylated to pyruvate by

pyruvate kinase (see Figure 44). This gives a cycle with a net conversion of 1 NADPH to 1 NADH. Unusual cycles like these have in recent years been found to operate in a number of Gram positive organisms. For example Sauer *et al.* (1997) reported a futile cycle in *Bacillus subtilis* involving phosphoenolpyruvate carboxykinase. Christiansen *et al.* (2001) have shown the cycle outlined above to be active at very low levels in slow growing *Bacillus clausii*. Similar cycles have also been observed in *Corynebacterium glutamicum* (Petersen *et al.* (2000). These groups used labeling techniques to determine the flux through these cycles. It is impossible to find these fluxes using flux balance analysis because inclusion of more than 1 anaplerotic pathway in a stoichiometric matrix for flux analysis yield a linear dependency making the fluxes impossible to determine. Without access to these techniques it was not possible to determine if this cycle occurred in *S. erythraea*. Another possible explaination for this is that a different isozyme of isocitrate dehydrogenase utilising NADH might be used in the stationary phase. However Roszkowski *et al.* (1971) and Alvarado and Flores (2003) found it to be strictly NADPH dependent.

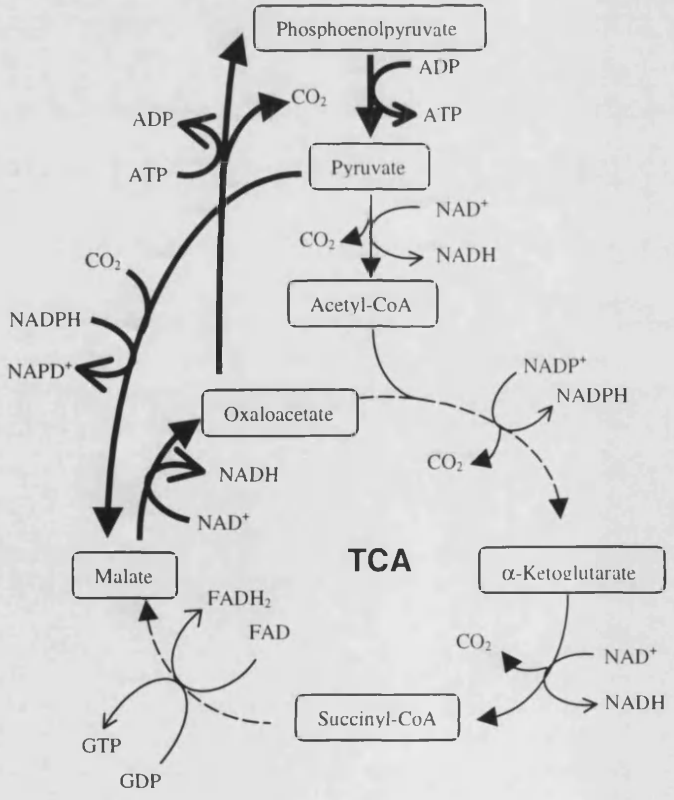


Figure 44: Metabolic cycle acting as a transhydrogenase

In the last few months Fischer and Sauer (2003) have published details of a pathway found to operate in *E. coli* under conditions when excess NADPH production occurs. The pathway operates in a cyclic manner analogous to the TCA cycle using many of the TCA cycle components but bypasses isocitrate dehydrogenase which is NADPH dependent. The pathway uses the glyoxylate shunt coupled with phosphoenolpyruvate carboxykinase to form a rather complex cycle. This pathway was published after the modelling in this project was completed. Inclusion of these pathways would in any case have caused a linear dependency in the stoichiometric matrix yielding it unsolvable.

In order to get around the problem of negative stationary phase fluxes in the PPP isocitrate dehydrogenase was considered to be NADH dependent. Conceptually this is rather similar to the pathway found by Fischer and Sauer (2003). Rather than relying on transhydrogenase activity which cannot be modelled it is assumed that the cell is capable of oxidising glucose fully without producing NADPH. The main differences between the results calculated in this way and those likely to have been seen with the pathway of Fischer and Sauer would be that. 1) Rather than cycling round the TCA cycle metabolites would be moving across it via the glyoxylate bypass. 2) The excess anaplerotic flux that the glyoxylate shunt generates would have to be converted back to phosphoenolpyruvate via phosphoenolpyruvate carboxykinase in order to enter the cycle once more. 3) The secretion of α-ketoglutarate would generate low levels of NADPH reducing the flux seen in the pentose phosphate pathway even further. Interestingly the fluxes seen in the PPP generating NADPH are of similar size to the fluxes to α-ketoglutarate which would generate NADPH in the scenario suggested.

4.3.4.2 Evaluation of Carbon Uptake and Carbon Leakage During Growth and Stationary Phase

4.3.4.2.1 Introduction

The change in the balance between the PPP and the TCA cycle is not the only large difference between the flux distribution seen in the growth phase and that seen in the stationary phase. There is also a shift from using carbon for growth to storing carbon and secreting carbon as organic acids. This seemingly wasteful use of carbon is investigated in this section.

4.3.4.2.2 Organic Acid Production

The impact that secreting organic acids might be expected to have on central metabolism is large. However the impact is actually much smaller than might be expected because the acids produced are all central metabolites and are secreted in a way that helps maintain the flux distribution in central metabolism. As the cell enters the stationary phase and growth stops there is no need for the synthesis of large amounts of biomass components. The fall in demand for biomass components derived from TCA cycle intermediates would be expected to lead to a fall in the flux through the anaplerotic reaction catalysed by phosphoenolpyruvate carboxylase. However the production of α-ketoglutarate and fumarate mean that compared with the glucose uptake anaplerotic flux is proportionately larger at 40 hours than at 25 hours. This maintenance of the flux distribution points towards leakage rather than the synthesis of organic acids for a specific purpose. Similar observations have been made with related organisms. Dekleva and Strohl (1988b) reported that the specific activity of phosphoenolpyruvate carboxylase increased when *Streptomyces* C5 moved from growth phase to production phase. In an analogous way pyruvate is one of the main sources of precursors for biomass in the exponential phase. In the stationary phase the cell maintains a high flux from this node but now going to organic acid secretion rather than biomass formation.

Throughout the growth phase central metabolism was supplying large fluxes to the precursors of biomass. This is done through key intermediates such as pyruvate, α-ketoglutarate and glucose-6-phosphate. Once growth stops a number of these points start to "leak" through secretion of metabolites or through production of storage compounds. It is as though *S. erythraea* is unable to adjust its uptake of glucose to meet its requirements. This leads to overflow of metabolites at points in metabolism that are set up to supply large amounts of these metabolites to biomass formation in the growth phase.

4.3.4.2.3 Control of Glucose Uptake

It is possible to investigate this idea further. An organism which tailored its uptake of glucose in the stationary phase to meet its needs would be expected to have an uptake

rate that was independent of the glucose concentration. It would only be expected to show limitation due to glucose concentration at low levels of glucose.

Figure 45 shows how glucose uptake varies with glucose concentration in *S. erythraea*. The transition from growth phase to stationary phase is visible as a clear kink at around 120 mmole/l. During the stationary phase glucose uptake seems to be dependent on glucose concentration. This is indicated by the approximately linear relationship until the glucose runs out. Although the uptake seems to be dependent on glucose concentration, it is not solely determined by glucose concentration or the line would be expected to pass through the origin. This suggests that *S. erythraea* displays only a limited amount of control over the rate at which it consumes glucose.

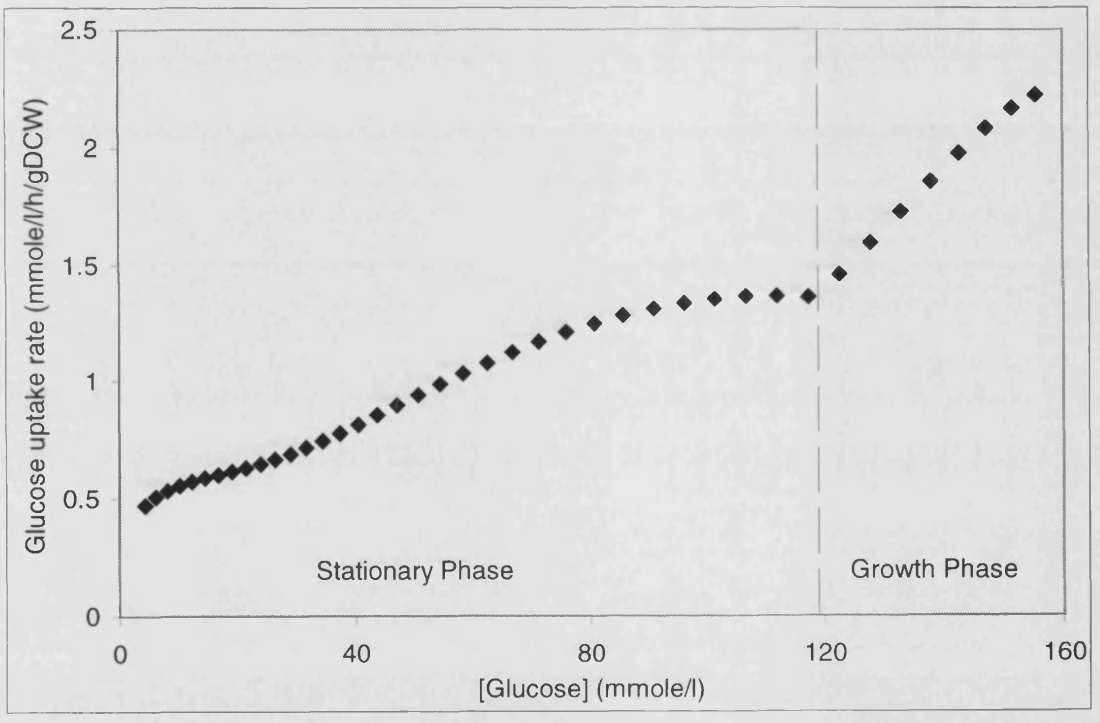


Figure 45: Comparison of how glucose uptake rate varies with concentration in white variant wild type fermentation 4.

This limited ability of the wild type organism to control its use of the carbon source according to its need may be a consequence of the grossly unnatural conditions it is being grown under. In the wild *S. erythraea* would be growing in the soil, a less rich environment, utilising slowly metabolisable carbon sources and with no mixing to aid mass transfer. Under those conditions it is unlikely that there would be a need to limit uptake of carbon source. In its natural environment *S. erythraea* may be in competition with faster growing soil bacteria able to utilise a simple substrate like glucose very

quickly. It may therefore be to *S. erythraea*'s advantage to take up glucose more quickly than it can use it in these circumstances in order to store it or convert it to a less readily useable carbon source such as organic acids. These can then be used later once the glucose has run out. Whilst an inability to control carbon uptake might explain why accumulation and secretion of carbon take place it does not explain why these particular metabolites are accumulated or secreted.

4.3.4.2.4 Reasons for Accumulation and Secretion of Certain Compounds

Considering the storage compounds first. The two storage compounds found in *S. erythraea* (trehalose and lipid) were both being synthesised in the growth phase. During the growth phase however their synthesis was balanced with cell growth preventing their accumulation in the biomass. Merely as a result of continuing to be synthesised when growth ceases they will accumulate in the biomass which has stopped increasing. The accumulation of trehalose and lipid in the biomass is then not particularly surprising as neither neutral lipids of trehalose contain nitrogen, the limiting nutrient.

The major organic acids produced by *S. erythraea* in the stationary phase are α-ketoglutarate and pyruvate. Both of these are precursors of dehydrogenase enzymes acting in central metabolism. Fumarate, the third organic acid produced, is under physiological conditions in equilibrium with malate which is the precursor of malate dehydrogenase. This may not be a coincidence as the rate of all these enzymes can be limited by the demand for NAD^+ and as has been shown by Hockenhull *et al.* (1954). In fact they found that in *Streptomyces griseus* inhibition of these enzymes leads to secretion of α-ketoglutarate and pyruvate. As the cell moves from growth phase to stationary phase the energy requirement for synthesis of biomass precursors drops. This leads to a fall in the rate of oxidative phosphorylation and thereby in the demand for NADH. At the same time demand for precursors of biomass also falls leading to more carbon being processed by the network. This fall in the demand for NADH would lead to accumulation of NADH reducing the intracellular levels of NAD^+. This will limit the flux through the TCA cycle and combined with poor control of glucose uptake may explain the link between secretion products and dehydrogenase enzymes. This hypothesis is explored further in Chapter 6. A similar idea has been proposed by

Surowitz and Pfister (1985). They found that in *Streptomyces alboniger* the glycolytic enzymes leading to pyruvate had higher activities when grown on glucose than the enzymes leading from pyruvate into the TCA cycle. They ascribed the secretion of pyruvate by this organism to the imbalance between glycolysis and the TCA cycle.

4.3.4.2.5 Summary

The second phase of the fermentation is characterised by production of organic acids and carbon storage compounds. This is thought to be due to an inability of the organism to exert control over the uptake of glucose to meet its requirements. The build up of the specific organic acids in question may be linked to the limited demand for dehydrogenase activity needed to supply the energy requirements of the organism in the stationary phase.

4.3.4.3 Analysis of Energy Demand During Growth and Stationary Phase

4.3.4.3.1 Introduction

Carbon and redox balances have provided insight into changes occurring as the cell moves from growth phase to stationary phase. There is another balance however which is intimately linked to the previous two, the energy balance. In order to remain healthy cells have to keep their ratio of ADP to ATP within certain limits. Under conditions of carbon excess with no growth such as occurs in the stationary phase the demand for ATP limits the rate of utilisation of carbon and of production of NADH. In this section the demand for energy within the cell will be investigated.

4.3.4.3.2 Energy Demand

The OUR measurements allow the rate of oxidative phosphorylation to be calculated in the cell. This must match the demand for ATP in the cell in order to maintain the ATP:ADP ratio. However the demand for ATP within the metabolic network does not match that calculated from the OUR. This is because there are a number of processes that use ATP which are not considered in the metabolic network such as protein turnover and maintenance of the membrane potential. These are termed the

maintenance energy requirements. These demands are not expected to vary greatly with the growth rate of the organism and so would be expected to remain fairly constant throughout the different stages of the fermentation.

The excess amount of ATP made in the network surplus to the requirements of the metabolic network is shown on the flux diagrams as a flux of ATP to maintenance. The excess use of ATP calculated is identical at 25 and 40 hours. This seems to be in line with the expectation of a constant maintenance requirement. However the value does not remain constant over the whole of the stationary phase. In fact the maintenance requirement drops from around 16 mmole ATP/gDCW/h at 40 hours to 11 ATP/gDCW/h at 50 hours and 8.5 ATP/gDCW/h at 60 hours. Moreover when the glucose is depleted it falls further so that at 67 hours it is 3.4 ATP/gDCW/h and at 90 hours 2.3 ATP/gDCW/h. It seems then that the maintenance requirements of *S. erythraea* vary according to its nutritional environment. Alternatively under rich nutrient conditions large amounts of energy available from central metabolism may not used for essential maintenance or for growth but instead are somehow discarded.

Literature values for maintenance energy requirements seem to vary greatly from 0-19 mmole/l/h (Stephanopolous *et al.* 1998, page 74). These results are not however in terms of mmole/gDCW/h therefore comparison is difficult. All the theoretical estimates calculated here fall within this range. Estimates for closely related organisms also vary considerably. Naeimpoor and Mavituna (2000) estimated maintenance energy requirements in an underdetermined metabolic network for *S. coelicolor.* Under nitrogen limitation with maintenance energy as the maximised objective function a maintenance energy of 10.39 mmole ATP/gDCW/h was found and a growth rate of 0.06 h^{-1}. With different nutrient limitations they found different maintenance energy requirements between 10 and 32 mmole ATP/gDCW/h. Roubos *et al.* (2001) estimated the maintenance energy requirements of *S. clavuligerus* grown under phosphate limitation to be 0.029 mol O_2/Cmol DCW/h. Using the elemental biomass composition for *S. clavuligerus* given in Roubos (2002) and the P to O ratio used in this work this gives a maintenance energy requirement of 4.2 mmole ATP/g DCW/h. Beyond noting that the maintenance energies calculated here lie within the range of those reported in the literature it is difficult to comment on how the values here fit in as there seems to be such large changes with the conditions the cells experience. When glucose is not

limiting the values found are of the same magnitude as the values found by Naeimpoor and Mavituna. When glucose is limiting the values found are closer to that found by Roubos *et al.* (2001). Note the comarisons are dependent on the value of the P:O ratio which is not known with any degree of certainty for these organisms.

The high fluxes through oxidative phosphorylation when glucose is in excess suggest that *S. erythraea* may be wasting energy. If this is the case then it raises some questions. Why are organic acids produced in the stationary phase when this carbon could be burnt in an analogous manner? Why is the wastage so high in the growth phase when nitrate is not limiting growth? What is the mechanism by which this energy is lost? It looks once more as if the cell is unable to control its uptake of glucose under the unusual condition applied in batch culture and wastage of energy is just another mechanism by which to accommodate this.

4.3.4.3.3 Summary

Under conditions of carbon excess the energy utilisation of *S. erythraea* for purposes not required by the metabolic processes considered changes hugely. This seems to point to energy being wasted both in the growth and stationary phase. This wastage of energy may well be linked to the wastage of carbon. Both may be due to limited control of carbon uptake.

4.3.4.4 Evaluation of the Late Phases of Fermentation

4.3.4.4.1 Introduction

The late phases of fermentation occur when glucose, the primary carbon source runs out. Although the glucose has been depleted there are still a couple of other carbon sources available to the cell. The storage compounds and organic acids produced in the previous phase serve as carbon sources. This leads to radically different flux distributions in the network yielding some more interesting insights as will be discussed later.

4.3.4.4.2 Flux Distributions Calculated after Glucose Depletion

Flux distributions for 67 and 90 hours have been calculated and are displayed as Figure 41 and Figure 42 respectively. These flux distributions have much in common. They are dominated by the fact that glucose has run out and the cell has to derive its carbon from other sources. Two of the key sources are fumarate and α-ketoglutarate which feed back into the TCA cycle. This creates the opposite problem to that seen when the cells are growing. When the cells are growing depletion of the TCA cycle intermediates requires replenishment by anaplerotic reactions to maintain the activity of the cycle. However when the carbon source enters via the TCA cycle the cycle is unable to run on its own as no acetyl-CoA is entering to allow the cycle to be completed. To get round this the phosphoenolpyruvate carboxykinase is used to regenerate phosphoenolpyruvate which then goes on to provide the acetyl-CoA so that the cycle can be completed. The flux through phosphoenolpyruvate carboxykinase reflects the flux of metabolites into the TCA cycle. It stays reasonably constant between the two time points despite the fact that fumarate is a major contributor at 67 hours but has been exhausted by 90 hours. The uptake of α-ketoglutarate seems to be increased to maintain the flux of acetate into the TCA cycle. This flux ensures the production of energy for cell maintenance.

The flux through glycolysis is greatly reduced after glucose depletion with intracellular stores of trehalose supplying the glycolytic flux that does occur. By 90 hours gluconeogenesis may be occurring although the fluxes are not large enough to rule out the possibility of measurement error causing this.

4.3.4.4.3 Summary

Flux in all metabolic pathways is greatly reduced after glucose depletion. The shape of the flux distribution is different as α-ketoglutarate becomes the main carbon source and carbon flows back up through phosphoenolpyruvate in order to maintain TCA cycle activity.

4.3.4.5 Classification of Nodal Rigidity

4.3.4.5.1 Introduction

The different flux distributions seen through the different stages of the batch fermentation may provide insight into how control is or is not exerted over the flux distribution by the regulatory mechanisms in the network. Branch points in the metabolic network can be classified according to how flexible they are with respect to the flux splits obtained at them (see section 1.4.3). The rigidity of some of the key branch points in the metabolic network can be found from the flux distributions calculated.

4.3.4.5.2 Rigidity of the Phosphoenolpyruvate Node

Consider phosphoenolpyruvate, at this node the network branches between continuance of glycolysis and anaplerotic feeding of the TCA cycle. If high yields of erythromycin are to be synthesised via succinyl-CoA in the TCA cycle it will be necessary to be able to control the flux through phosphoenolpyruvate carboxylase to maintain precursor availability. Figure 37 to Figure 42 show that the flux distribution around phosphoenolpyruvate varies enormously at different stages in the fermentation. Fluxes from oxaloacetate to pyruvate use phosphoenolpyruvate carboxykinase rather than phosphoenolpyruvate carboxylase so the flux through the latter enzyme in these cases must be regarded as zero rather than negative. The flux distribution at this node is not rigidly controlled. That is to say the flux through pyruvate kinase is not linked to the flux through phosphoenolpyruvate carboxylase or visa versa.

4.3.4.5.3 Rigidity of the Glucose-6-Phosphate Node

Likewise the branch point at glucose-6-phosphate which is crucial for supplying the deoxysugars attached to erythromycin shows a high degree of flexibility. Fluxes from glucose-6-phosphate can go into glycolysis, into the pentose phosphate pathway or into glucose-1-phosphate synthesis. The simplified flux diagrams presented hide the fact that the sugars for erythromycin biosynthesis branch off from the sugars for biomass formation at glucose-1-phosphate rather than glucose-6-phosphate. Between 25 and 40 hours the flux to glucose-1-phosphate remains virtually unchanged whilst the fluxes to

fructose-6-phosphate and 6-phosphogluconolactone change hugely. This suggests that the glucose-6-phosphate node is flexible and should be amenable to metabolic engineering techniques.

4.3.4.5.4 Potential Problems of Applying Rigid Node Analysis

There is a potential shortcoming with this form of analysis for an organism such as *S. erythraea*. If the *S. coelicolor* genome is representative of actinomycetes *S. erythraea* probably contains multiple isozymes of many of the important genes of central metabolism. These isozymes would allow the organism to use enzymes with different control properties at different stages of the fermentation thereby preventing the use of this type of analysis. For example *S. coelicolor* has two pyruvate kinase genes (Hesketh *et al.* 2002). It is entirely feasible that one might be used in the growth phase and the other in the stationary phase. In fact only one of the glycolytic genes in *S. coelicolor* does not have an isozyme (Hesketh *et al.* 2002).

4.3.4.5.5 Summary

Important nodes in the central metabolism of *S. erythraea* seem to be flexible. This means that they ought to be amenable to metabolic engineering attempts to alter the flux split at these points. It is not surprising that organisms that show great metabolic flexibility should have flexible nodes in their metabolic networks. There is however a chance that this apparent flexibility might not be actual flexibility but rather be "hard wired" by the used of different isozymes at different stages of the fermentation.

4.3.5 Comparison with Results from Recent Publications

During the course of this project a number of papers have been published by groups working to use similar techniques to investigate flux distributions in *streptomyces* species. A quick overview of their findings and comparison with the findings in this project is given here.

Daae and Ison (1999) performed flux balance analysis on the central metabolism of *Streptomyces lividans* during the growth phase. They found that the pentose phosphate pathway displayed very high fluxes. The same was found to be true of *S. erythraea* in this work.

Naeimpoor and Mavituna (2000) produced an FBA model of *Streptomyces coelicolor* including the pathways for production of actinorhodin, a polyketide. The fermentation was performed in a chemostat at a dilution rate of 0.06 h^{-1}. This makes the results rather difficult to compare with the work presented here as a growth rate of 0.06 h^{-1} is characteristic of neither exponential growth or stationary phase in this organism. In their case the flux distributions show no flux through the oxidative part of the PPP. This stands in marked contrast to the large PPP fluxes found when *S. erythraea* was growing exponentially. The main aim of their paper was to calculate theoretical maximum growth rates and maintenance energy requirements. However given that the fermentation was performed at a dilution rate of 0.06 h^{-1} their conclusion that the maximum growth rate achievable at steady state is 0.093 h^{-1} seems rather questionable. What the work does highlight is that under nutrient limitation when the carbon source is in excess the utilisation of the carbon source is not closely linked to the growth rate. So rather than maximising growth rate in an under determined system better results were found by maximising maintenance energy. This suggests that the organism is wasting energy under these conditions. This finding is in close agreement with my own.

Kirk *et al.* (2000) performed flux balance analysis on *Streptomyces clavuligerus* under carbon, nitrogen and phosphorus limitation. The experiments were performed in chemostat culture at a dilution rate of 0.05h^{-1}. The use of a chemostat at an intermediate dilution rate once again limits the comparability of their results with the work done here. *S. clavuligerus* showed different production rates for clavulanic acid under different nutrient limitations. The flux distributions are somewhat different from those found in this work. The PPP does not play a dominant role even though there is a significant growth rate. The difference in the importance the PPP has in Kirk *et al.* (2000) and in Naeimpoor and Mavituna (2000) compared with this work is probably down to the choice of nitrogen source. Nitrate was used in this work and it must be considerably reduced before use, this requires NADPH. The lower requirements for NADPH can be supplied by TCA cycle activity alone. Kirk *et al.* (2000) found very

large fluxes in the urea cycle under during growth, this is probably because the urea cycle seems to be involved in nitrogen assimilation. With *S. erythraea* fluxes in the urea cycle were very small, it was limited to synthesis and catabolism of arginine. The main conclusion of the work is that for the production of a compound containing carbon and nitrogen in appreciable amounts, phosphate limitation is a better choice than carbon or nitrogen limitation.

Jonsbu *et al.* (2001) measured fluxes in *Streptomyces noursei* in batch culture using C13 labelling. This method is not effected by transhydrogenase activity so the results should be an accurate reflection of what is happening in the PPP. Flux distributions were found for several time points through the fermentation allowing the differences between the growth phase and the stationary phase to be studied. Unfortunately the synthesis of nystatin in *S. noursei* does not display a classic secondary metabolite profile. Although production of nystanin does not start until 25-30 hours into the fermentation there is still growth occurring and when growth stops nystatin production also stops. This makes comparison between *S. noursei* and *S. erythraea* rather difficult. Although growth is slower during the production phase it is still present and has a dominating effect on some of the fluxes. For example the flux through the PPP falls as the production phase is entered but it falls much less than in *S. erythraea* as there is still substantial cell growth. As the flux through the PPP falls the relative flux through the TCA cycle increases. This general pattern fits with what is seen in *S. erythraea.* One difference between *S. noursei* and *S. erythraea* is that as *S. noursei* enters production phase the anaplerotic flux falls as the growth rate decreases. In *S. erythraea* this flux goes up as large amounts of α-ketoglutarate start to be synthesised. *S. noursei* seems to show closer control over its uptake and utilisation of glucose than *S. erythraea* does.

Avignone Rossa *et al.* (2002) performed flux balance analysis of *Streptomyces lividans* in chemostat culture under phosphate limitation. They used three strains, a wild type and two overproducing strains, one for actinorhodin (ACT) and one for undecylprodigiosin (RED). Two carbon sources were used: glucose and gluconate. Gluconate feeds into the pentose phosphate pathway instead of into glycolysis forcing a very different flux distribution from the chemically similar glucose. The first dehydrogenase enzyme in the pathway is also avoided leading to no NADPH being

formed when gluconate enters the PPP as the second dehydrogenase is NADH dependant. They found that at higher growth rates use of the PPP relative to the TCA cycle went up and antibiotic production went down regardless of carbon source. This is in line with the results found here where the flux through the PPP is high during exponential growth and low during the stationary phase. Avignone Rossa *et al.* also found that growth on gluconate produced much higher levels of organic acid secretion than growth on glucose. This is interesting because by using gluconate as carbon source the ratio of NADPH to NADH produced in the PPP drops. In order to achieve the same growth rate more NADH has to be synthesised. This indicates a tight link between organic acid overproduction and NADH synthesis similar to the one proposed here for *S. erythraea* in section 4.3.4.2.4.

Avignone Rossa *et al.* (2002) also found that overproducing strains produced more of their respective secondary metabolites than the wildtype strain. The levels however were not high enough to make a significant impact on the metabolic network. Secondary metabolite production even in these overproducing strains did not put a significant metabolic burden on the cell compared to biomass production. From the results they were able to make a testable prediction about the production of ACT and RED. They predicted that decreased pentose phosphate flux would lead to lower biomass formation and higher ACT and RED yields. This is counter intuitive as both antibiotics require NADPH for their synthesis. The predictive power is illustrated by the fact that Butler *et al.* (2002) independently reduced flux through the PPP in *S. lividans*. They expected it to decrease antibiotic production and found that instead it was increased. Sequeria *et al.* (2003) have now characterised the flux distributions in *S. coelicolor* strains with reduced PPP activity and have confirmed the earlier results. This illustrates the point that whilst producing very high yields of a product may require one set of changes to the organism's metabolism modest changes can be achieved by completely different approaches.

Roubos (2002) has presented the most comprehensive FBA model to date for a *streptomyces* species using *S. clavuligerus*. Unfortunately his results are not particularly comparable with those found here. This is because *S. clavuligerus* does not use glucose as a carbon source, so a range of other carbon sources were used. Of those used maltose was the closest to glucose. In addition the nitrogen source used was

glutamate and *S. clavuligerus* often had a higher uptake rate for glutamate than for the carbon source. This leads to a large proportion of the carbon entering metabolism through the TCA cycle, yielding very different flux distributions. Roubos' emphasis is on changes of flux distribution in the growth phase caused by using different carbon sources rather than the difference between growth phase and stationary phase. Whilst the TCA cycle fluxes and the anaplerotic fluxes are not comparable the growth phase PPP flux is comparable. The flux through the PPP in the growth phase is much lower in *S. clavuligerus* than in *S. erythraea.* Partly this may be due to the use of nitrate as nitrogen source in this work because nitrate requires reduction to ammonia before it can be utilised and this conversion requires NADPH.

Flux balance analysis has also been used to study penicillin production in *Penicillium chrysogenum* (Jørgensen *et al.* 1995, van Gulik *et al.* 2000). The work of van Gulik is particularly interesting as the flux balance analysis is used to determine the rigidity of the principle nodes of the network. They found that none of the nodes were rigid and hence the network had plenty of flexibility to increase penicillin production. This is interesting as the same situation is found is *S. erythraea.* It could be a general principle in organisms performing secondary metabolism that flexibility is required in order to allow the metabolic flux distribution to adapt as nutrient conditions change. van Gulik *et al.* also used different carbon and nitrogen sources to manipulate the amount of available NADPH in the cell and found that this did seem to have an impact on penicillin production. It appeared that cofactor availability was the key factor controlling metabolism in this network. Cofactor availability may have been more important in their work because a high yielding strain was used whereas it was not so important in the *streptomyces* work discussed above because the yield of those strains were too low to make a significant impact on NADPH demand.

Summary

Although a number of papers applying flux balance analysis to organisms similar to *S. erythraea* have been published during the course of this project none of them are directly comparable due either to different growth conditions or different patterns of growth. However despite this some of similarities emerge, high PPP flux during growth is one such feature.

4.3.6 Comparison of Theoretical and Actual Flux Distributions

4.3.6.1 Introduction

Elementary mode analysis is a technique for identifying all the possible truly unique pathways that can be formed in a metabolic network. This set of different modes of operation can be linearly combined to form any flux distribution possible in the organism (see section 1.4.4). From the set of elementary modes it is possible to find the modes with the highest yield of product on substrate (Klamt *et al.* 2003). This has been done for a subset of the network composed for *S. erythraea* in section 2.10. The flux distributions achieving these theoretical yields will be compared with the flux distributions found in the wild type during the stationary phase.

The elementary modes with the highest erythromycin to glucose ratio using protein and RNA as nitrogen sources are presented as Figure 46 and Figure 47 respectively. The mode using protein gave a yield of 0.415 g erythromycin/g glucose. The mode using RNA gave 0.399 g erythromycin/g glucose. Both modes share some interesting characteristics in common.

4.3.6.2 Description of the Elementary Modes Calculated

Both modes utilise a reversed TCA cycle to deliver carbon to propionyl-CoA and methylmalonyl-CoA, the precursors of the polyketide moiety of erythromycin. This method gives a higher yield than operating the TCA cycle in the normal direction. This is the case because the carbon would in any case have to enter the TCA cycle via anaplerosis ie via phosphoenolpyruvate carboxylase. To then convert this to succinyl-CoA via the usual direction of the TCA cycle would require one pyruvate molecule to be decarboxylated to acetate. This acetate molecule would then be incorporated into citric acid and 2 molecules of CO_2 removed before succinyl-CoA is formed. This would greatly reduce the conversion efficiency.

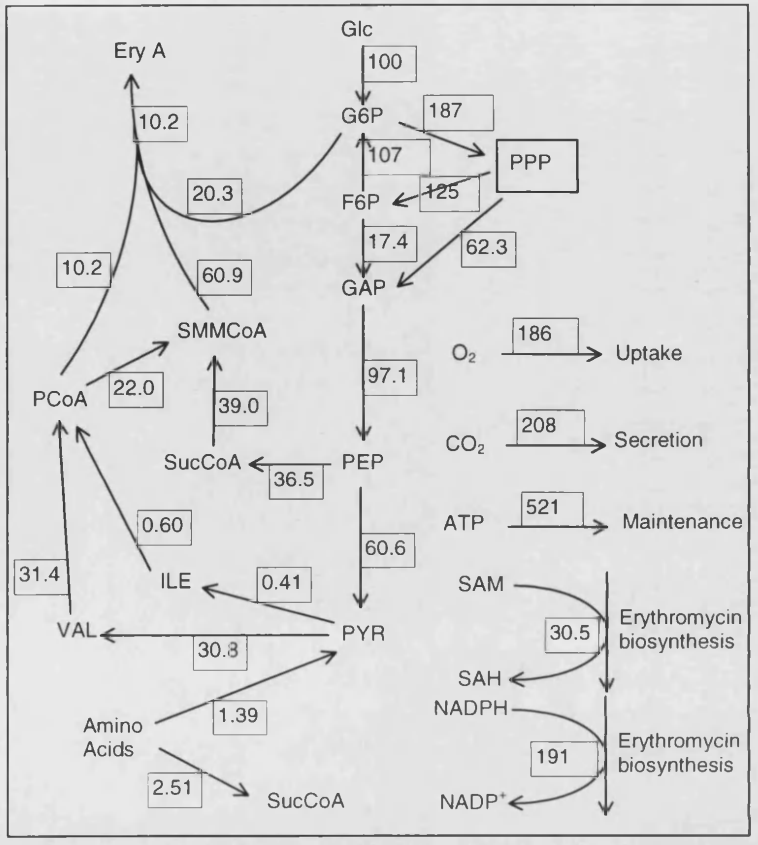


Figure 46: Elementary mode giving maximum theoretical yield of erythromycin on glucose and protein for the metabolic network of *S. erythraea*. The maximum theoretical yield is 0.102 $mole_{erythromycin}/mole_{glucose}$ or 0.415 $g_{erythromycin}/g_{glucose}$. This yield was achieved by two theoretical yields, the one above and a similar mode in which all flux from pyruvate to product occurred via valine rather than via isoleucine and valine. Whilst the majority of the flux from phosphoenolpyruvate to succinyl-CoA proceeds via malate in a reverse of the usual TCA direction a small proportion proceeds via citrate in order to provide a route to catabolise acetyl-CoA formed by the degradation of some amino acids.

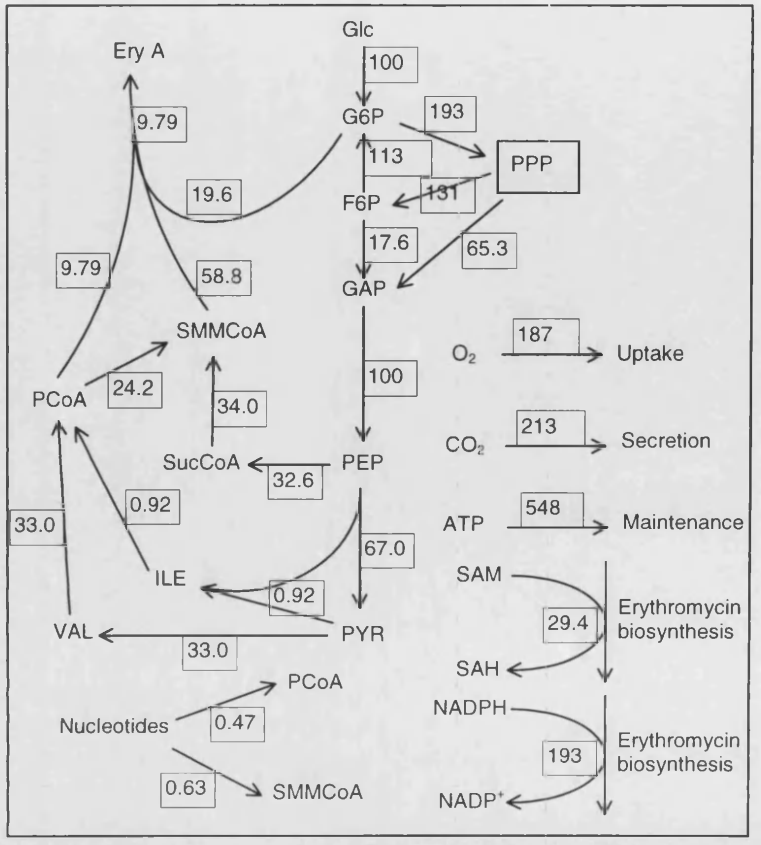


Figure 47: Elementary mode giving maximum theoretical yield of erythromycin on glucose and RNA for the metabolic network of *S. erythraea*. The maximum theoretical yield is 0.098 $mole_{erythromycin}/mole_{glucose}$ or 0.399 $g_{erythromycin}/g_{glucose}$. This yield was achieved by two theoretical yields, the one above and a similar mode in which all flux from pyruvate to product occurred via valine rather than via isoleucine and valine.

In the case where protein is supplying the nitrogen required for erythromycin synthesis there is some flux through the TCA cycle in the normal direction. This occurs because acetyl-CoA is produced in the catabolism of some amino acids and further catabolism of this acetyl-CoA requires the operation of the TCA cycle in the normal direction. In this case however the carbon lost in the TCA cycle to CO_2 is as much as that acquired from the amino acid derived acetyl-CoA. The conversion of glucose to methylmalonyl-CoA is just as efficient as for a reversed TCA cycle.

Both modes use more than one pathway to deliver precursors to polyketide synthesis. The mode using protein as nitrogen source uses a split TCA cycle as the primary source of precursors. It also synthesises extra valine and isoleucine so that they can be broken down to supply propionyl-CoA. The mode using RNA as nitrogen source also uses the split TCA cycle as primary source of precursors, it also synthesises valine in order to catabolise it to propionyl-CoA.

The mixed utilisation of carbon sources is a consequence of the requirement for the cell to balance cofactor requirements. The most efficient way to make methylmalonyl-CoA from glucose is via a reversed TCA cycle because this requires half a glucose per methylmalonyl-CoA formed. The formation of methylmalonyl-CoA from a normal use of the TCA cycle or from valine or isoleucine requires a whole glucose molecule. The reversal of the TCA cycle however consumes $FADH_2$ forming FAD^+, the $FADH_2$ must be regenerated if the pathway is to be sustainable. The production of propionyl-CoA from valine or isoleucine regenerates $FADH_2$ from FAD^+ via the enzyme acyl-CoA dehydrogenase. By combining routes via valine and isoleucine with the route via the reversed TCA cycle in roughly equal proportions the cell is able to balance its cofactor requirements whilst giving the highest sustainable yield of erythromycin possible.

4.3.6.3 Comparison of Flux Distribution for Maximum Erythromycin Yield with Flux Distribution from the Stationary Phase Wild Type Organism

These elementary modes demonstrate how complex and integrated the pathways needed to synthesise erythromycin at theoretical yields are. The flux distributions depicted here are radically different from those found in the stationary phase metabolism of the wild type organism. Comparison will be made with the stationary phase because such high

yields are incompatible with significant growth as this would compete for the precursors being produced by central metabolism. In order to achieve theoretical yields in *S. erythraea* large changes would have to be made to its metabolism in a number of places and ways. Some of these changes are considered below.

Evidence from *S. erythraea* suggests that in the wild type organism it is the pathways of erythromycin biosynthesis that limit its overproduction (see Rowe *et al.* 1998). Clearly all the pathways specific to erythromycin biosynthesis would have to be up regulated including the PKS enzymes, the sugar manipulation pathways and the enzymes for post PKS modifications. It might also be necessary to up regulate or improve self-protection mechanisms to prevent auto-toxicity.

Not surprisingly elementary modes analysis reveals that PPP activity needs to be high in order to support NADPH requirements for polyketide synthesis. In the wild type organism PPP activity is low in the stationary phase because there is little demand for NADPH. It may be that *S. erythraea* is able to increase flux through this pathway to meet demand. The glucose-6-phosphate node certainly seems to be flexible. However it could be that capacity in the PPP is reduced in response to low demand for NADPH in the stationary phase. Given the apparent flexibility of the glucose-6-phosphate node increasing capacity should be sufficient to overcome limitations on the availability of NADPH.

The unusual operation of the TCA cycle required to achieve theoretical yields shows interesting similarities but also differences when compared to what is actually seen in the wild type organism. In the wild type organism during stationary phase large amounts of α-ketoglutarate are produced. This requires high levels of anaplerotic flux. The best yields of erythromycin also require high anaplerotic fluxes, considerably higher than seen in the wild type. The analysis in section 4.3.4.5.2 seems to show that the phosphoenolpyruvate node is flexible. Increasing the capacity of phosphoenolpyruvate carboxylase and the demand for succinyl-CoA from the TCA cycle should result in the desired flux. After entering the TCA cycle the carbon is then processed in the opposite direction to that normally seen in *S. erythraea*. The process is certainly thermodynamically feasible and succinate fermentation using a split TCA

cycle is well known in some groups of fermentative bacteria (Stephanopoulos *et al.* 1998, Page 87). However to my knowledge it has never been observed in actinomycetes. This unusual flux might be achieved by restricting or removing an enzyme such as citrate lyase in order to force the carbon round the TCA cycle in the opposite direction. It would remain to be seen how amenable the TCA cycle enzymes in *S. erythraea* are to this.

High levels of erythromycin synthesis require high levels of methylation. *s*-Adenosylmethionine methylation certainly occurs in stationary phase *S. erythraea.* However capacity would probably need to be increased to meet higher demand. The supply of amino acids required to balance $FADH_2$ demand would also have to be engineered. These pathways are already active in the stationary phase supplying short branched chain organic acids for iso and anteiso fatty acid synthesis. The capacity however would have to be increased. It might be possible to feed these amino acids in the broth. This however might have the adverse effect of making more NH_3 which could cause growth rather than secondary metabolism. Secondly the long term feeding of antibiotics at a large scale is likely to be economically undesirable compare with synthesis of the same amino acids in the host.

The supply of deoxysugars requires higher fluxes to glucose-1-phosphate than seen in the wild type in stationary phase. However given that the pathways are already active and that the glucose-6-phosphate node seems to be flexible this would probably only require the up regulation of phosphoglucomutase. However the synthesis of trehalose or polysaccharides from this glucose-1-phosphate would have to be prevented.

The fluxes considered above would require up-regulation. However other fluxes would have to be down regulated to prevent wasteful synthesis of side products. Some fluxes such as that to trehalose could be blocked by deletions because they do not play an important role in the required flux distribution. Others such as the flux to pyruvate are required indirectly for the synthesis of erythromycin. In these cases a careful balance would have to be achieved to control synthesis to only that level which is required.

4.3.6.4 Summary

Flux distributions giving the maximum possible yields of erythromycin from glucose and biomass components broken down in the stationary phase have been identified. These are radically different from the yields found in the wild type organism requiring a split TCA cycle. Large changes to the metabolism of *S. erythraea* would be necessary to implement this flux distribution.

4.4 Conclusions

Batch fermentation in *S. erythraea* can be divided into three main phases. During these phases the patterns of metabolism are very different owing to different driving forces and this showed in the flux distributions found. The growth phase is dominated by the requirements of supplying cofactors and precursors for growth this leads to high PPP and anaplerotic activity along with high levels of biomass synthesis.

The stationary phase prior to glucose depletion is characterised by low PPP fluxes and low production of precursors and cofactors for biomass synthesis. Instead large scale leakage of carbon and energy occurs. The dominating factor in this phase of the fermentation is the glucose uptake rate. *S. erythraea* does not seem able to control the rate of glucose uptake. This leads to the wasteful use of carbon and energy. *S. erythraea*'s inability to control glucose uptake may be due to the fact that in its natural environment it is unlikely to be exposed to such high levels of glucose. The wastage of glucose and energy slowly declines as the glucose is depleted. Erythromycin synthesis occurs in this phase.

Once glucose is fully depleted another metabolic paradigm ensues. The rate of respiration falls sharply, showing that the earlier high levels were greater than necessary to maintain the cell's basal activities. The flux distribution is dominated by the reuse of α-ketoglutarate and trehalose synthesised in the previous phase to generate energy via the TCA cycle. The use of α-ketoglutarate as carbon source causes reversal of anaplerosis giving a very different flux distribution. There is no erythromycin synthesis during this phase.

During the course of a batch fermentation *S. erythraea* exhibits great metabolic flexibility. This flexibility is shown in the ability to change flux split ratio at key branch points in the metabolic network such as phosphoenolpyruvate and glucose-6-phosphate. This indicates that the key nodes in the network are not rigidly controlled. Networks displaying this kind of flexibility are more easily manipulated by metabolic engineering approaches.

Elementary modes analysis was performed on the metabolic network of *S. erythraea.* From these modes the flux distribution with the highest yield of erythromycin on glucose and protein was found. The maximum yield required a split TCA cycle. This distribution is radically different from that seen in the wild type organism during the stationary phase. Many changes would need to be made to the organism in order to implement this distribution.

The application of flux balance analysis to batch fermentations of *S. erythraea* has allowed some of the key properties of the different phases to be identified. A deeper understanding of the metabolic processes involved has led to new hypotheses regarding the importance of the uptake of glucose during the stationary phase. It has also led to new hypotheses regarding the relationship between oxidative phosphorylation, the TCA cycle and organic acid secretion during the stationary phase. In the next chapter an industrial strain of *S. erythraea* will be compared with the parental strain. These four areas of metabolism i.e. glucose uptake, oxidative phosphorylation, TCA cycle activity and organic acid secretion will feature prominently in understanding the differences between the strains. In chapter 6 the relationship between the TCA cycle, oxidative phosphorylation and organic acid secretion will be investigated in more detail using an uncoupler of oxidative phosphorylation.

5 COMPARISON OF DIFFERENT STRAINS

5.1 Introduction

In the last chapter the differences seen in the metabolism of *S. erythraea* between different phases of a batch fermentation were investigated. A number of key factors were found which it was postulated were responsible for some of the unusual behaviour seen in this organism in the stationary phase such as organic acid secretion. In this chapter the wild type white variant is compared with an industrial strain screened for improved erythromycin production and the wild type red variant strain. The key areas of metabolism identified in the last chapter are considered in the comparison of the strains and are found to show considerable differences between the strains.

One approach to identifying strategies for increasing productivity of microbial products is to study overproducing strains constructed using traditional strain improvement techniques to find the causes of their increased productivity. This strategy is known as reverse metabolic engineering and has been discussed in section 1.5.2.3. An industrial strain of *S. erythraea* (CA340) which overproduces erythromycin was available. Fermentation and flux balance analysis were performed with this strain to identify differences between CA340 and the parental white variant.

Khosla and Bailey (1988) applied the reverse metabolic engineering approach to *S. erythraea.* They introduced a haemoglobin gene from another organism into a industrial strain of *S. erythraea* to improve oxygen transfer and found erythromycin production was improved. However to the best of my knowledge reverse metabolic engineering has not been used to compare an industrial strain with its wild type parent for secondary metabolite production in actinomycetes before. A brief assessment of the utility of the method for this kind of product and organism has been included.

A comparison has also been made between the red variant and the white variant in order to try to assess some of the differences between these two wild type strains. This is

important because historically industrial overproducing strains have been based on the white variant eg strain CA340 (Seno and Hutchinson 1986). However most recent genetic engineering of *S. erythraea* has been performed in the red variant because it is easier to manipulate (P. Leadley, personal communication).

Based on the comparison of these strains a strategy that may be generally applicable to the overproduction of secondary metabolites in batch cultures of actinomycetes is identified. The potential for improvement using this strategy is considered along with the implications it might have on further improvement of a hypothetical strain.

5.2 Comparison of *S. erythraea* CA340 and the White Variant

5.2.1 Introduction

S. erythraea CA340 is an industrial strain of *S. erythraea* developed by Abbot laboratories, the strain produces 10 times more erythromycin than the parental strain (Reeves *et al.* 1998). It was produced by a classical strain improvement program hence the nature of the changes made to the strain are unknown. Restriction endonuclease studies suggest that the changes made are not extensive, those differences found using this method were not in region in which the genes for erythromycin synthesis are found (Reeves *et al.* 1988). Elucidating the nature of some of these changes by comparison of flux distributions with the wild type may make it possible to introduce beneficial changes into other related organisms directly. Given that the changes are not extensive (Reeves *et al.* 1998) it is considered acceptable to use the same metabolic network with this strain as the wild type.

This approach presented some difficulties. CA340 has been selected for production in complex media. Flux analysis depends on accurate measurements of the rate of utilisation of media components which is not possible with complex media, therefore analysis has to be performed in defined media. However fermentation in defined media does not truly reflect the situation in complex media. This restricts the method to showing general changes in the overall metabolism rather than the specifics of overproduction in the industrial situation. Understanding these general changes is

however key to developing strains whose general metabolism is capable of supporting a large overproduction of a side product distant from central metabolism.

This section aims to identify key differences between the flux distributions seen in strain CA340 and in the parental strain and to consider how these differences might effect the yields of erythromycin produced by the strains.

5.2.2 Materials and Methods

S. erythraea CA340 was grown in duplicate bioreactors under identical conditions to those used in section 4.2 for the white variant wild type. All biochemical analysis was performed using the same protocols. The modelling techniques used were also identical to those described in chapter 4.2.

5.2.3 Results and Discussion

5.2.3.1 Comparison of Fermentation Profiles Between the Parental Strain and CA340

In order to determine growth and product kinetics dry cell weight, glucose, nitrate, OUR, CER, α-ketoglutarate, pyruvate, fumarate, extracellular protein, erythromycin, red pigment, total organic carbon, intracellular protein, RNA, DNA, trehalose, arabinogalactan and lipid were measured over the course of 160 hours. The summed sources of carbon in the broth agreed well with the total organic carbon analysis suggesting all important sources were being measured. The nitrogen balance was not as complete. In the parental strain on average 85% of nitrogen was accounted for during the growth phase, whereas in the stationary phase large amounts of nitrogen were not accounted for (see section 4.3.2). In strain CA340 the nitrogen balance in the exponential phase was somewhat worse, only 75% of the nitrogen was accounted for. The stationary phase nitrogen balance was similar to that seen in the parental strain. For this reason care was taken in the use of the nitrate uptake rate during the flux balance analysis. The impact of this rate on the flux distribution was assessed on a case by case basis for the exponential phase and if necessary this rate was not used as an input for the

model. Since the model is overdetermined the rate is calculated based on the other data. During the stationary phase a similar strategy was employed and a theoretical flux for secretion of the excess nitrogen was provided (see section 2.10.7). This was done in order reduce the adverse effect of the incomplete nitrogen balance on the calculation of the other rates.

The fermentation was performed in duplicate, the two fermentations agreed closely. For the sake of brevity only one set of results is presented. The fermentation profiles measured are presented with their counterparts for the parental strain in Figure 49 to Figure 53. There are some interesting differences between the fermentation profiles displayed by strain CA340 and those seen with the parental strain. Firstly Strain CA340 grows more slowly than the parental strain with a μ_{max} of 0.043 h^{-1} instead of 0.088 h^{-1}. This is reflected in all the profiles which spread over a longer fermentation time. Some of the other key differences are examined below.

The specific glucose consumption rate is slower in the overproducing strain than in the parental strain. Moreover the glucose is not used up by the end of the fermentation. As a consequence of this the off gas profiles are somewhat different in CA340. There is no sudden drop in the OUR and CER marking the depletion of glucose as is seen in the wild type. The off gas profiles also differ in the relative size of the peak OUR and CER readings. In the parental strain the peak level of the CER in the stationary phase is significantly lower than the peak level in the growth phase. Whereas in CA340 the peak level in the stationary phase is of a similar magnitude to the peak level in the growth phase. Furthermore the time difference between the two peaks is also smaller in CA340. The same pattern is seen in the OUR readings.

The amounts of α-ketoglutarate, pyruvate and fumarate secreted by CA340 are lower than those secreted by the parental strain. What is more the α-ketoglutarate and fumarate do not seem to be as readily taken back into the cell either, presumably because glucose does not run out so there is no need to remetabolise them. The levels of intracellular trehalose show a similar trend both strains make similar amounts however CA340 does not reconsume its trehalose unlike the parental strain.

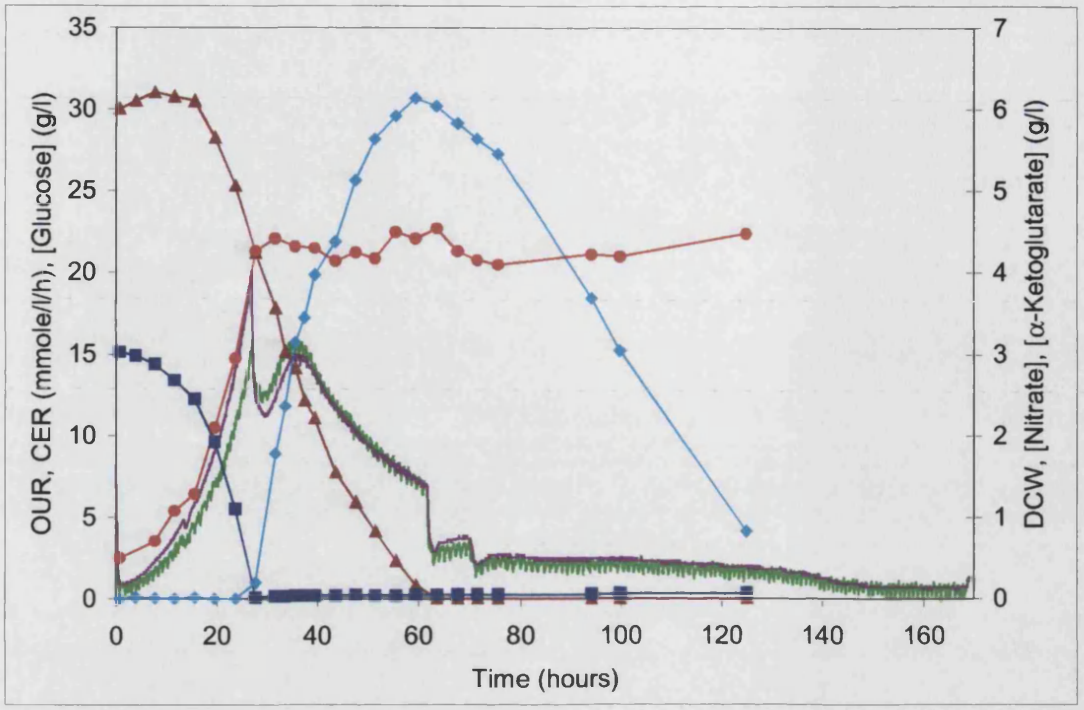


Figure 48: White variant wild type fermentation 5 profile 1. OUR (mmole/l/h) — , CER (mmole/l/h) — , Glucose (g/l) ▲ , Nitrate (g/l) ■ , α-Ketoglutarate mono sodium salt (g/l) ◆ , Dry cell weight (g/l) ● .

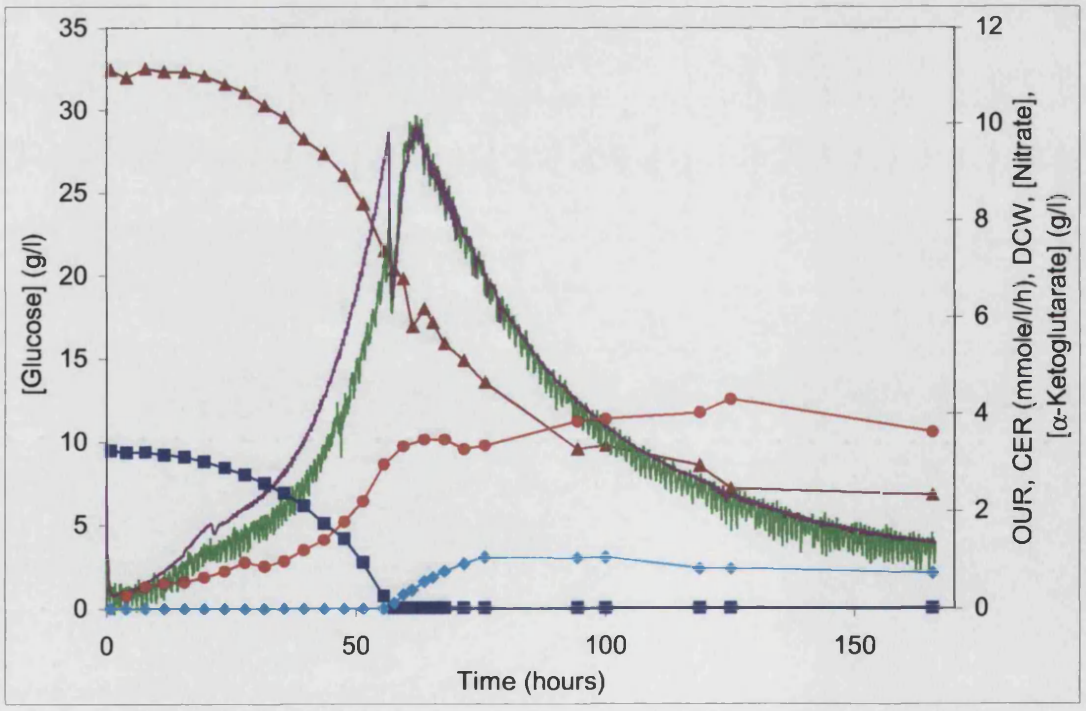


Figure 49: Strain CA340 fermentation 1, profile 1. OUR (mmole/l/h) — , CER (mmole/l/h) — , Glucose (g/l) ▲ , Nitrate (g/l) ■ , α-Ketoglutarate mono sodium salt (g/l) ◆ , Dry cell weight (g/l) ● .

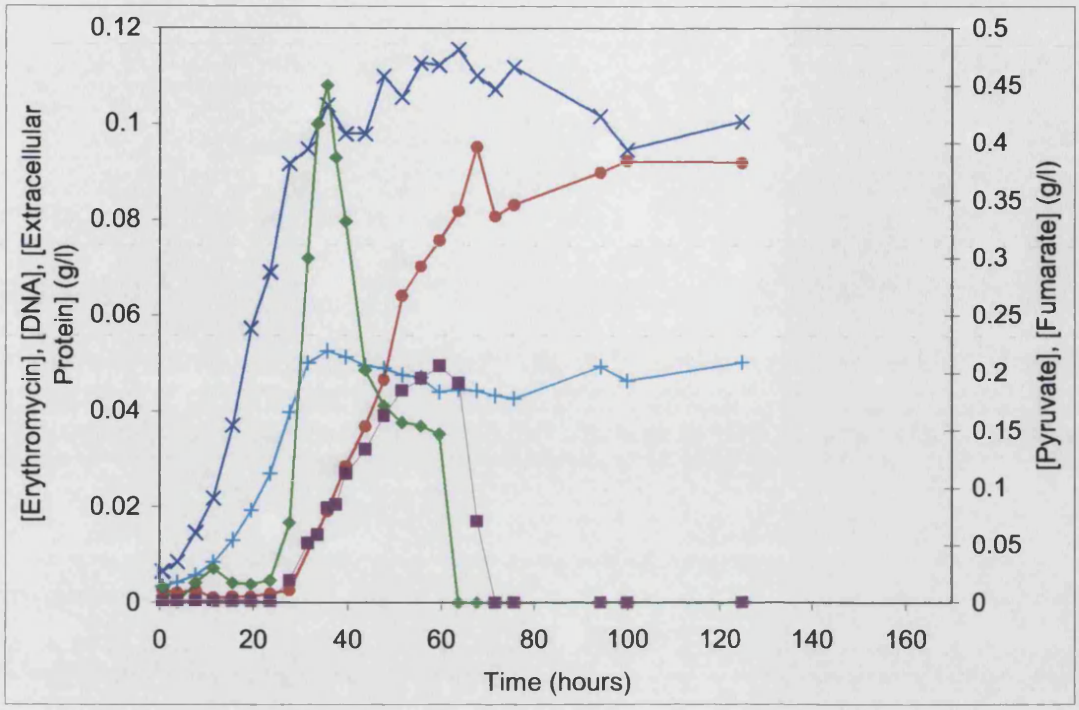


Figure 50: White variant wild type fermentation 5 profile 2. Fumarate disodium salt (g/l) ■ , Extra cellular protein (g/l) × , Erythromycin (g/l) ● , DNA (g/l) + , Pyruvate (g/l) ◆ .

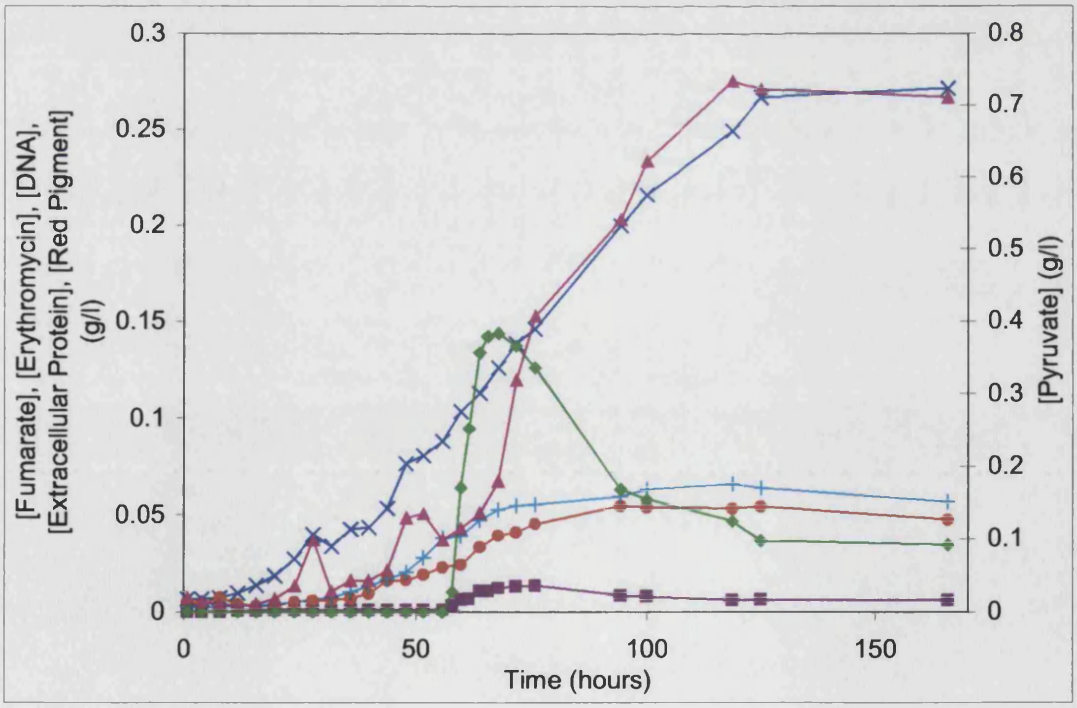


Figure 51: Strain CA340 fermentation 1, profile 2. Fumarate disodium salt (g/l) ■ , Extra cellular protein (g/l) × , Erythromycin (g/l) ● , Red pigment (g/l) ▲ , DNA (g/l) + , Pyruvate (g/l) ◆ .

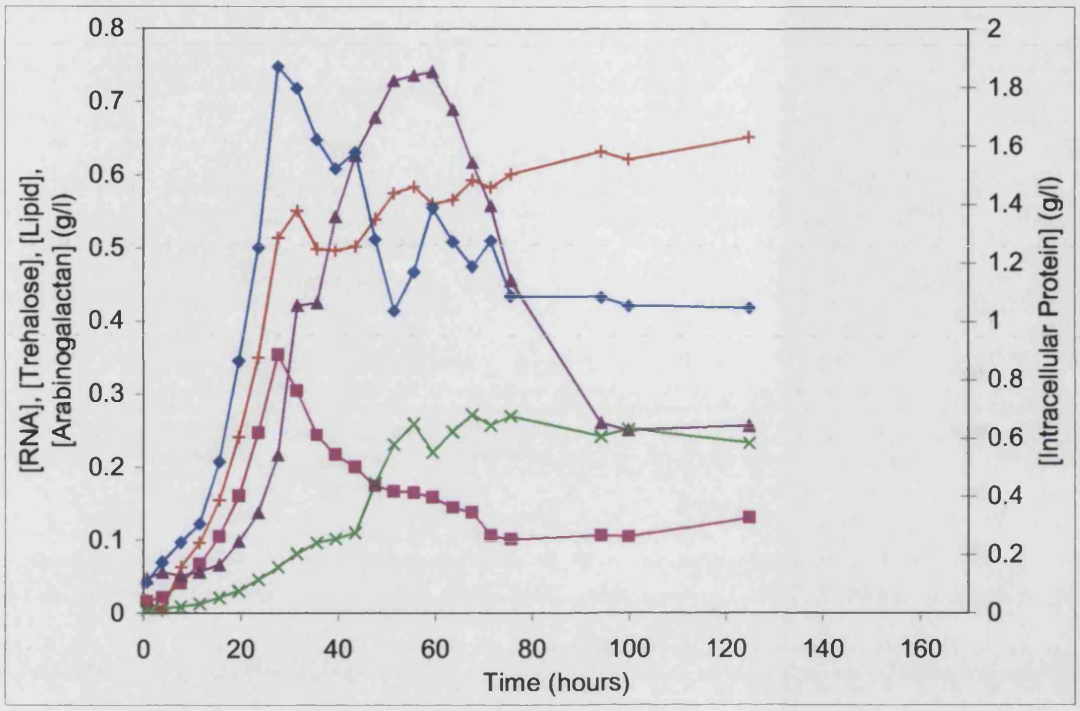


Figure 52: White variant wild type fermentation 5 profile 3. RNA (g/l) ■ , Lipid (g/l) × , Trehalose (g/l) ▲ , Arabinogalactan (g/l) + , intracellular protein (g/l) ◆ .

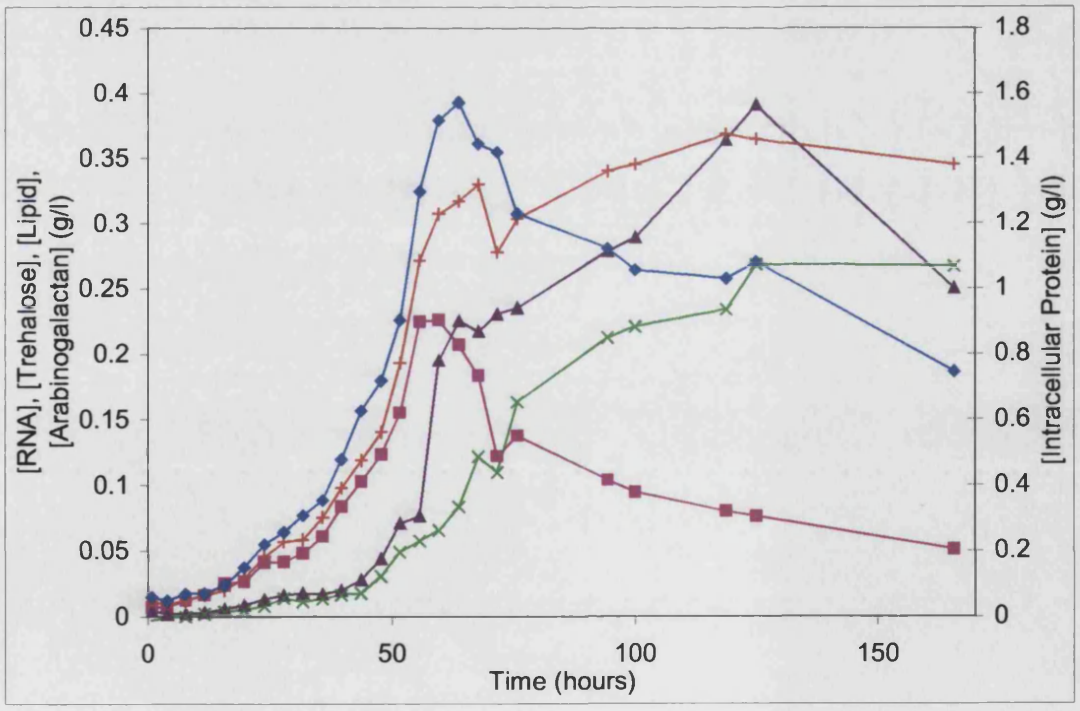


Figure 53: Strain CA340 fermentation 1, reactor 1, profile 3. RNA (g/l) ■ , Lipid (g/l) × , Trehalose (g/l) ▲ , Arabinogalactan (g/l) + , intracellular protein (g/l) ◆ .

In defined media strain CA340 does not produce as much erythromycin as the wild type strain (overproduction is still seen on complex media). It does however produce large amounts of the red pigment. The production of secondary metabolites appears not to be as tightly linked to the stationary phase as in the parental strain. The production of both erythromycin and red pigment is seen in the exponential phase. The synthesis of these compounds carries on in the stationary phase.

The protein, arabinogalactan and lipid levels and their profiles are broadly similar between the two strains. The RNA levels are lower in CA340 this is probably due to the lower growth rate. Shahab *et al.* (1996) found that RNA levels rose as growth rate increased in chemostat studies of *S. coelicolor.*

5.2.3.2 Comparison of Flux Distributions Between the Parental Strain and CA340

The biomass profiles were used to calculate the uptake rates of the nutrients and production rates of the products and biomass components. These were then used to calculate flux distributions for a range of time points throughout the fermentation. Summaries of two such flux distributions, one from the growth phase and one from the stationary phase of CA340 can be found as Figure 54 and Figure 55. The corresponding figures for the parental strain are figures Figure 37 & Figure 38 in section 4.3.4. The figures are a reduced representation of the metabolic network used. They contain a simplified representation of central metabolism and an indication of most of the fluxes through which carbon leaves or enters central metabolism. Whilst some of the nodes may not seem to balance this is due to cycling of metabolites in a manner too complex for such a simple representation. A fuller, though still not complete, representation of the network used in the calculations was presented as Figure 7 (section 2.10).
When a flux is said to be undefined this means that the measured data for that flux has not been used in the model because its inclusion greatly worsens the fit achived with the model.

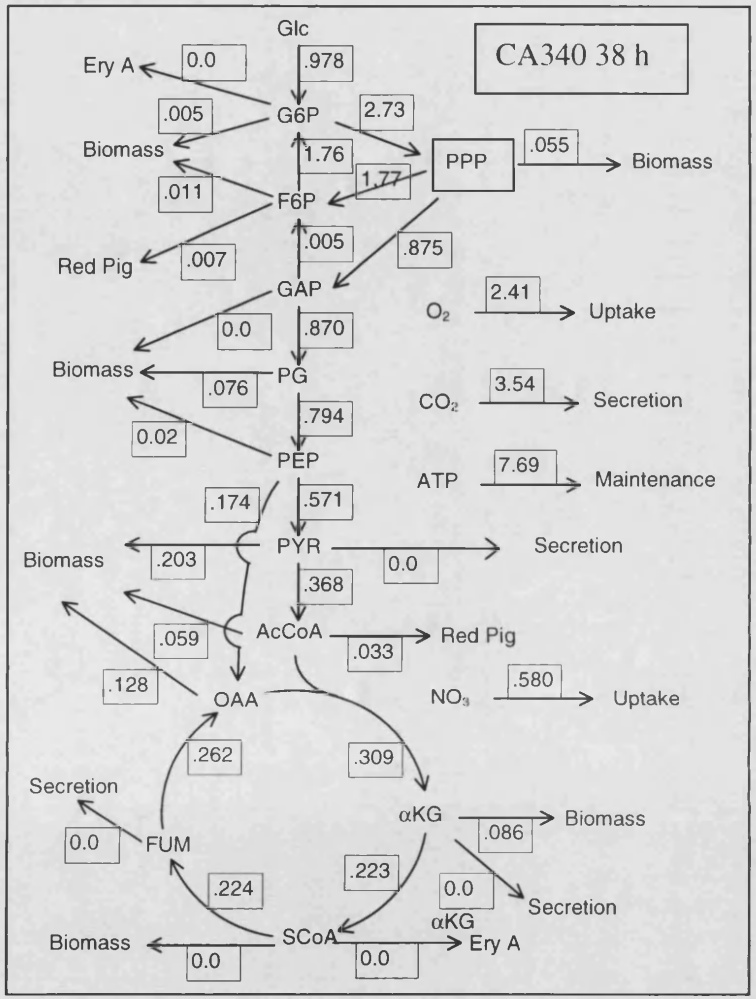


Figure 54: CA340 growth phase flux distribution. Flux to glucose undefined, weighted residual = 3.9. All fluxes are in mmole/gDCW/hour

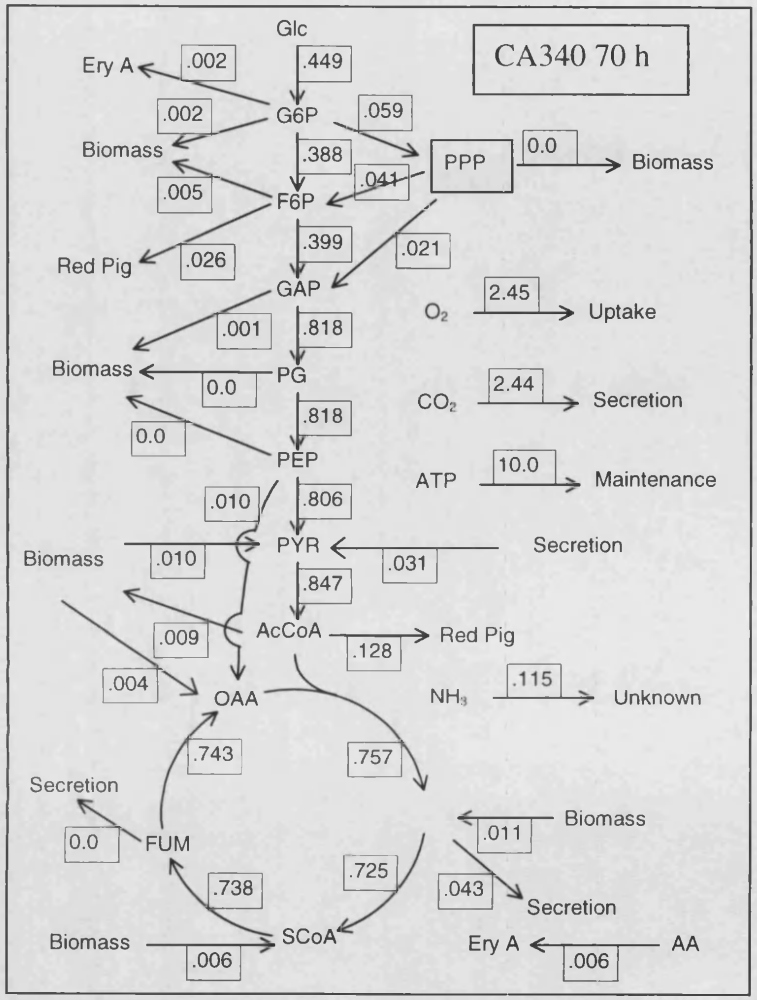


Figure 55: CA340 stationary phase flux distribution. Flux of ammonia from biomass undefined, weighted residual = 0.017. All fluxes are in mmole/gDCW/hour

The results from five of the key fluxes in the model have been presented graphically as Figure 56 to Figure 60. The pathways chosen are: Glucose uptake through hexokinase. PPP flux through glucose-6-phosphate dehydrogenase. TCA cycle flux through isocitrate dehydrogenase. Anaplerotic flux through phosphoenolpyruvate carboxylase. And the flux of ATP to unknown maintenance requirements. To aid comparison the graphs have been aligned. Zero hours is redefined as the time that nitrate becomes depleted. In this way comparison is facilitated because most of the large changes occur at this time.

The specific glucose uptake profile for CA340 is lower than that for the parental strain throughout the growth phase and for most of the stationary phase too until the parental strain runs out of glucose (see Figure 56). The parental strain uses glucose at a high rate until it runs out causing a precipitous drop in the OUR and CER values (see Figure 48). CA340 however uses glucose more slowly throughout the fermentation. In the stationary phase the glucose uptake rate falls more gently as the glucose concentration falls and the strain does not utilise all the glucose available.

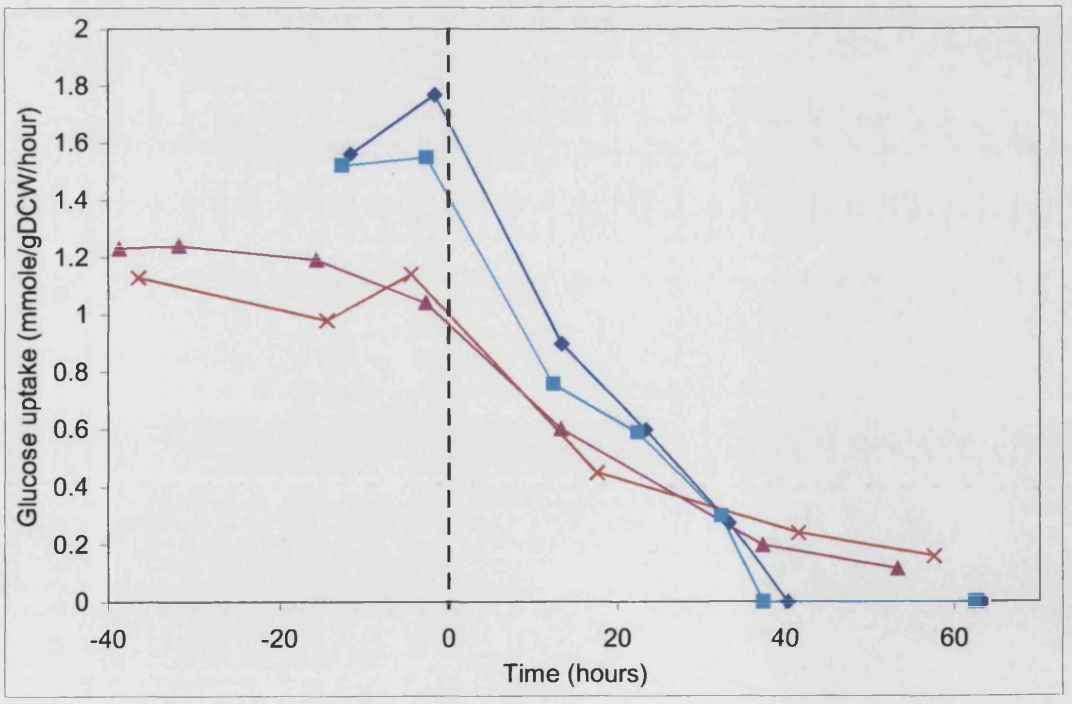


Figure 56: Comparison of glucose uptake rates between CA340 and the wild type strain. White variant wild type fermentation 4 ◆, White variant wild type fermentation 5 ■, CA340 fermentation 1 ▲, CA340 fermentation 2 ×. The profiles have been aligned by defining the depletion of nitrate as zero hours, this is marked by a dashed line.

The glucose uptake rate has been plotted against the residual glucose concentration in the medium for both strains in Figure 57. This plot highlights the differences in glucose utilisation between the two strains. The different phases can be identified by the differing rates and trends of glucose uptake. The exponential phase starts when the glucose concentration falls to about 120 mM. The wild type profile in the stationary phase seems approximately linear until the glucose falls to around 5 mM. A linear profile suggests that the mass transport of glucose is limiting glucose uptake. However this is unlikely to be the case for two reasons. Firstly if glucose were limiting the profile would be expected to pass through the origin, however it does not. Secondly the linear region seems to start at 90 mmole/l i.e. about 16 g/l. Glucose uptake is unlikely to be mass transport limited at such a high concentration. Rather another factor must be limiting uptake until the residual glucose concentration gets very low. At this point glucose mass transport becomes the limiting factor and the linear relationship is broken. The true limiting factor is probably the cell's ability to take up glucose. The uptake rate is slowly decreased during the stationary phase in a manner that looks similar to the glucose depletion yielding the apparently linear relationship in Figure 57.

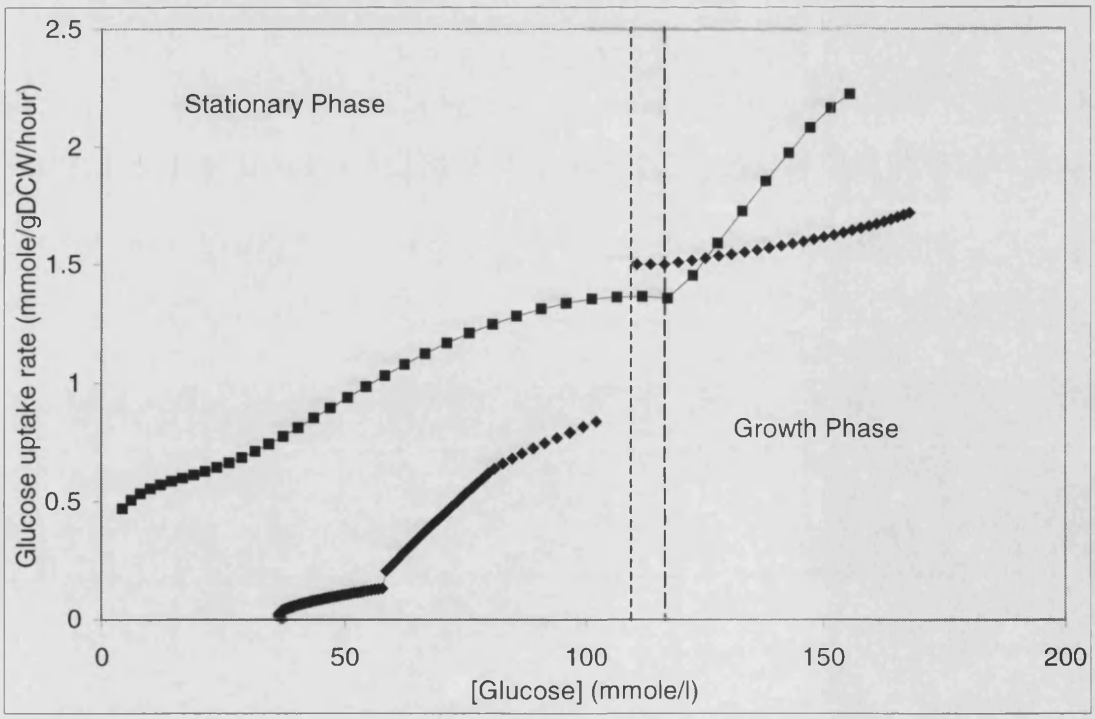


Figure 57: Variation of glucose uptake rate with glucose concentration for white variant wild type and CA340. White variant ■, CA340 ◆. Line with short dashes shows end of growth phase for CA340, line with long dashes shows end of growth phase for the parental strain.

The pattern for CA340 is somewhat different to that seen with the parental strain. The glucose uptake rate in the exponential phase is lower that seen in the parental strain. This phase however lasts longer and more glucose is consumed during it than in the parental strain. The cells enter stationary phase when the glucose drops to around 110 mM. At this point there is a sharp drop in the glucose uptake rate. The rate of utilisation is lower throughout the stationary phase than in the parental strain, it also falls more quickly with the glucose concentration and more or less stops before the glucose is depleted. The glucose uptake rate is not limited by the glucose concentration.

S. erythraea is a soil bacterium which naturally grows in an environment where there is intense competition for nutrients. In its natural environment it is unlikely to ever experience such high levels of glucose. In order to take advantage of any sugars available it may have high levels of constitutively expressed sugar uptake systems. This would allow it to scavenge any glucose available as quickly as possible. However because glucose would not normally be encountered in high concentrations there may be no regulation of these uptake systems. Such regulation would enable the organism to balance supply and demand for glucose. This could lead to the observed phenomena of organic acid secretion and wastage of energy seen in the parental strain when exposed to high levels of glucose (see sections 4.3.4.2.4 and 4.3.4.3). Strain CA340 has been improved by random mutagenesis combined with screening. It is possible that one of the glucose uptake systems in this strain has been inactivated or damaged reducing the glucose uptake rate. The industrial screening process has selected a strain that takes up glucose more slowly. Slow uptake is more appropriate for this kind of fermentation as less carbon is wasted in side products and unnecessary respiration.

The lack of control of the glucose uptake rate in the wild type leads to inefficient use of carbon as large quantities are taken up only to be secreted as organic acids. These organic acids are utilised again later in the fermentation when the glucose runs out. However this only occurs at a very slow rate which does not seem to be associated with erythromycin production (see section 4.3.4.4). By contrast in CA340 a slower uptake of glucose reduces this wasteful use of carbon.

The pentose phosphate flux in CA340 is broadly similar to that seen in the wild type organism during both phases (see Figure 58). This was unexpected because during the growth phase CA340 has a lower glucose uptake rate. This implies that a greater proportion of carbon is being directed towards NADPH production in the CA340 than in the white variant. CA340 however grows more slowly than the wild type. A slower growing organism would be expected to have a lower demand for NADPH than a faster growing organism because the biomass synthesis rate is reduced.

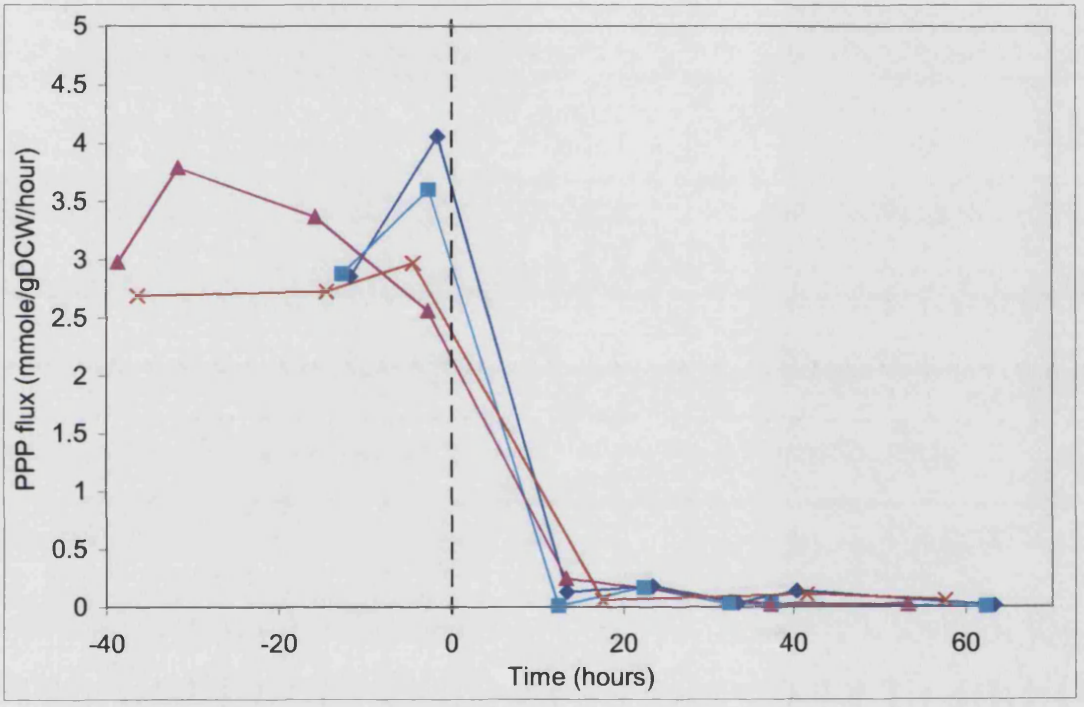


Figure 58: Comparison of calculated PPP flux between CA340 and the wild type strain. White variant wild type fermentation 4 ◆, White variant wild type fermentation 5 ■, CA340 fermentation 1 ▲, CA340 fermentation 2 ×. The profiles have been aligned by defining the depletion of nitrate as zero hours, this is marked by a dashed line.

CA340 grows more slowly than the parental strain. It would therefore be expected that CA340's maintenance energy requirements would account for a larger portion of its carbon utilisation than the parental strain and its growth a smaller portion. This is expected because the maintenance energy is associated with non-growth related upkeep of the cell. Examples of this include macromolecular turnover and dissipation of membrane potential. The cell's maintenance energy requirements are closely related to the rate of flux through the TCA cycle. It would be expected that maintenance energy levels (Figure 60) and TCA cycle fluxes (Figure 59) would be similar between the two strains whilst the PPP flux (Figure 58) would be lower in the slower growing strain. In

reality the TCA cycle flux and maintenance requirements are lower in CA340 and the PPP flux the same.

The maintenance energy requirements of CA340 (Figure 60) seem to be lower than in the wild type organism for most of the fermentation. This is unexpected as the maintenance energy should be approximately constant even between related strains. However as has been demonstrated in section 4.3.4.3 the level of unattributable energy usage in *S. erythraea* varies considerably even within a fermentation. This raises questions as to whether this is really maintenance associated energy. Nevertheless CA340 uses much lower amounts of ATP for maintenance compared with the wild type strain.

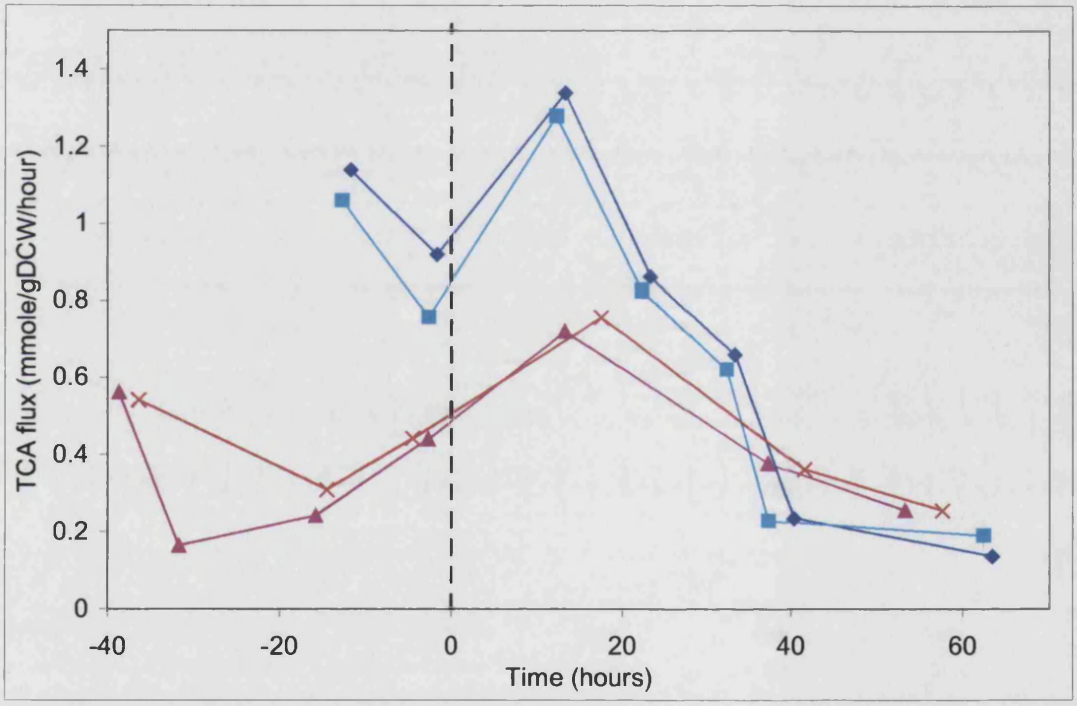


Figure 59: Comparison of calculated TCA cycle flux between CA340 and the wild type strain. White variant wild type fermentation 4 ◆, White variant wild type fermentation 5 ■, CA340 fermentation 1 ▲, CA340 fermentation 2 ×. The profiles have been aligned by defining the depletion of nitrate as zero hours, this is marked by a dashed line.

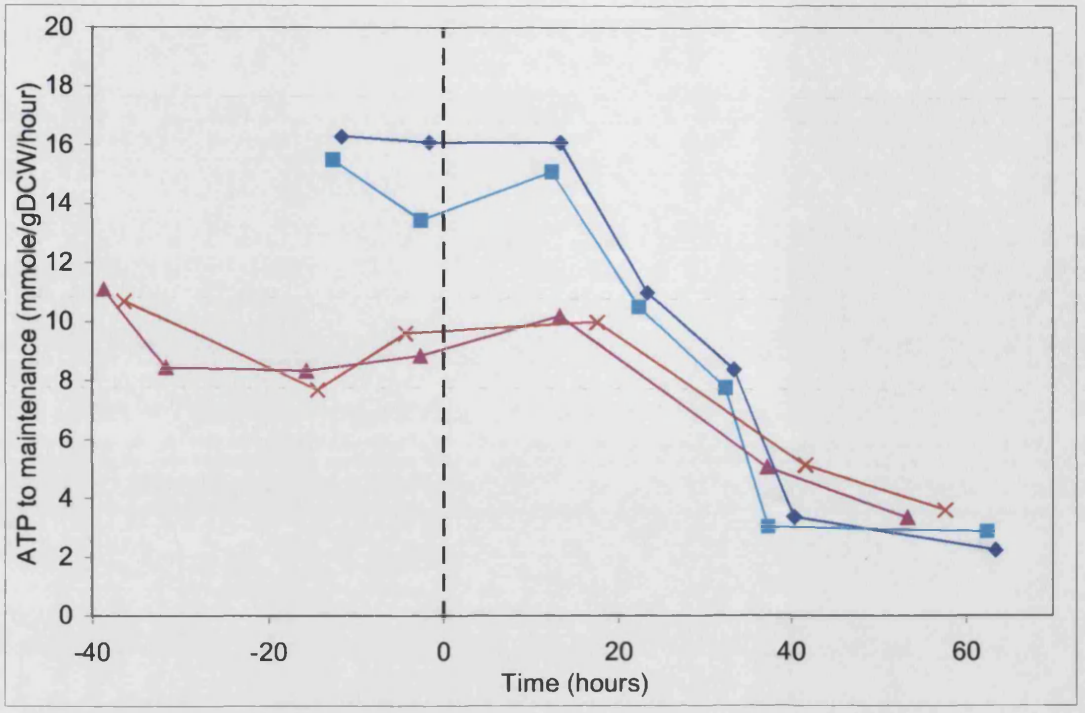


Figure 60: Comparison of calculated maintenance use of ATP between CA340 and the wild type strain. White variant wild type fermentation 4 ◆, White variant wild type fermentation 5 ■, CA340 fermentation 1 ▲, CA340 fermentation 2 ✕. The profiles have been aligned by defining the depletion of nitrate as zero hours, this is marked by a dashed line.

These observations regarding the use of the PPP and maintenance energy levels imply that CA340 is more efficient with its use of energy and carbon but more wasteful with its use of NADPH. These properties are both fitting for a strain selected for the overproduction of secondary metabolites. A lower maintenance energy requirement allows for a higher proportion of the carbon available to be converted to product. Overproduction of NADPH in the growth phase allows production of secondary metabolites in the growth phase thereby increasing the length of the production phase. In this case production of both measured secondary metabolites was seen in the exponential phase in CA340 but not in the parental strain (see Figure 50 and Figure 51). Under these conditions the differences did not lead to overproduction of erythromycin. It did however support overproduction of the red pigment, another polyketide, in comparison to the parental strain.

There is always a cause and effect problem with this type of analysis. The question is does a change in the cofactor ratios cause production of secondary metabolites or has the production of secondary metabolites caused central metabolism to shift to produce

an appropriate ratio of cofactors? Whichever is the case the CA340 strain is controlling its metabolism in a way that encourages overproduction of secondary metabolites and more efficient use of carbon.

The issues of the efficiency of use of carbon and the overproduction of NADPH have recently been investigated by Avignone Rossa *et al.* (2001). They found that decreasing NADPH production caused improved carbon efficiency and production of actinorhodin. In this case NADPH overproduction and high carbon efficiency were mutually exclusive. The case may be rather different for *S. erythraea* growing under these conditions because of the large amount of carbon wastage by other means such as excessive maintenance requirements and organic acid production. Loss of carbon through increased PPP operation could be made up by reducing waste in these areas.

The flux through phosphoenolpyruvate carboxylase (Figure 61) represents the amount of anaplerotic replenishment occurring in the cell. It replaces intermediates drained from the TCA cycle for growth or product formation. For example α-ketoglutarate and erythromycin both drain metabolites from the TCA cycle for their production. As expected the faster growing parental strain has the highest anaplerotic flux during the growth phase. The high production of α-ketoglutarate by the wild type in the first part of the stationary phase continues this trend until the glucose runs out (see Figure 48). In the wildtype once the glucose runs out the process is reversed. α-Ketoglutarate feeds into TCA cycle and back up through phosphoenolpyruvate carboxykinase in order to provide acetate for the TCA cycle flux. This gives a negative flux through phosphoenolpyruvate carboxylase, or more correctly a positive flux through phosphoenolpyruvate carboxykinase. This reversal of flux is not as apparent in strain CA340 because the external glucose supply does not run out and so α-ketoglutarate uptake is much lower. Rather in CA340 the phosphoenolpyruvate carboxylase flux settles to around zero towards the end of the fermentation as metabolic activity in the cell slows down.

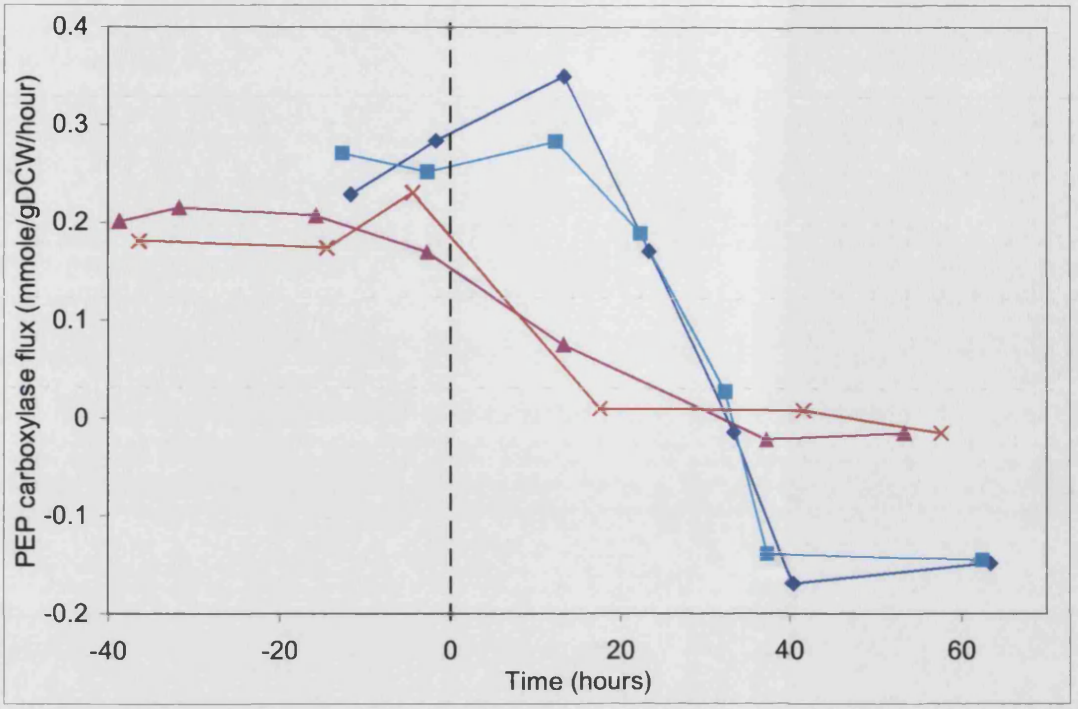


Figure 61: Comparison of calculated anaplerotic flux between CA340 and the wild type strain. White variant wild type fermentation 4 ◆, White variant wild type fermentation 5 ■, CA340 fermentation 1 ▲, CA340 fermentation 2 ×. The profiles have been aligned by defining the depletion of nitrate as zero hours, this is marked by a dashed line.

In CA340 the combination of slower glucose uptake, slower growth and increased secondary metabolite production during growth is interesting. When metabolism is limited by something other than carbon, organic acid secretion is seen as demonstrated by the wild type in the stationary phase. Therefore it seems reasonably to suggest that carbon is limiting the growth of CA340 during the growth phase as no organic acid secretion is seen. If the growth rate is being limited by the carbon uptake rate then the culture must be to some degree carbon limited throughout the growth phase. *S. erythraea* is known to produce erythromycin under carbon limitation (McDermott *et al.* 1993). This partial carbon limitation during the growth phase may explain the production of secondary metabolites in the growth phase.

Interestingly this analysis also gives a potential explanation for the nitrogen balance in the exponential phase being worse in CA340 than in the parental strain. Strain CA340 seems to display a number of properties usually associated with the stationary phase during the growth phase. In section 3.3.4 capsular material was found to be made by the parental strain during the stationary phase. It was proposed that if the capsule were

formed of polyglutamate this might account for the problems in closing the nitrogen balance during the stationary phase. CA340 displays some charicatristics of secondary metabolism in the expoential phase. If CA340 produces capsular material in the exponential phase then this might explain why the nitrogen balance is worse for strain CA340 than for the parental strain during the exponential phase. This could be investigated using the same methods as used in section 3.3.4.

Many of the properties that *S. erythraea* CA340 displays which are conducive to overproduction of secondary metabolites are potentially linked to one observable property of the organism, its low glucose uptake rate. Its better efficiency with carbon leading to a longer production phase, its slower growth rate that appears to be carbon limited and the production of erythromycin in the exponential phase. Industrial fermentations for antibiotic production often use slowly released carbohydrate sources such as starch for this purpose. It seems that strain selection for overproduction has acted to enhance this effect by producing a strain that utilises glucose slowly. This increases the potential for partial carbon limitation in the growth phase and therefore growth associated product formation. It also prevents wasteful production of organic acids and storage compounds in the stationary phase. It should be noted however that CA340 actually produces lower yields of erythromycin than the wild type under the conditions used in these studies.

Although the changes seen in strain CA340 lead to overproduction of erythromycin it should be noted that these changes are not particularly relevant for high levels of erythromycin overproduction. Firstly CA340 fails to use all the glucose available to it, it would be better if this carbon was also used for the synthesis of product. Furthermore in strain CA340 production has been increased by reducing inefficient wastage of carbon. The reduction of secretion of organic acids and excessive energy use increases the length of the production phase considerably. However if very high yields of erythromycin are required the organism will have to have a much higher uptake of glucose than needed to meet its maintenance energy demands. This carbon will have to be directed to erythromycin production. The very property that has been selected for in strain CA340 to increase erythromycin production may be incompatible with very high levels of overproduction. There is a danger that it will not be easy to develop strain CA340 further because it may have found a local optimum for erythromycin production.

If it has found a local optimum it may make it more difficult for this strain to achieve the global optimum than the wild type.

5.2.4 Summary

CA340 produced less erythromycin in this defined medium than the parental strain. However a number of the properties of strain CA340 are conducive to improved erythromycin production. CA340 utilises the available glucose more slowly. This reduces organic acid formation and maintenance energy levels leading to a longer stationary phase for erythromycin production. CA340 also produces erythromycin in the growth phase increasing the production phase still further. This too may be linked to the slow uptake of glucose as the strain may be partially glucose limited in the growth phase. These improvements however may not be conducive to further development of the strain as high productivity will probably require high glucose uptake rates.

5.3 Comparison of Red Variant and White Variant

5.3.1 Introduction

One of the interesting aspects of the microbiology of *S. erythraea* is the existence of two strains which display very distinct characteristics and yet interchange with a rather high frequency of 1-5% (Seno and Hutchinson 1986). These two strains are often referred to as the red variant and the white variant. The names are derived from pigments produced. The red variant produces large amounts of a soluble red pigment. The white variant produces a white spore pigment upon sporulation. There are other differences however, the red variant does not sporulate on many media, grows more slowly and produces less erythromycin. Strain CA340 is derived from the white variant and yet produces large quantities of the red pigment in the defined medium used in these studies. Comparing the red and the white variants may help identify other characteristics shared by the red variant and CA340. The comparison is also important because the white variant has been the strain of choice for industrial overproduction of erythromycin, however the red variant is the strain of choice for manipulation of the

genes of the PKS. If novel polyketide products are to be overproduced in *S. erythraea* a better understanding of the two wild type strains will be important.

5.3.2 Materials and Methods

S. erythraea red variant was grown in under identical conditions to those used in section 3.2 for the white variant wild type. All biochemical analysis was performed using the same protocols. The modelling techniques used were also identical.

5.3.3 Results and Discussion

5.3.3.1 Comparison of Fermentation Profiles Between White Variant and Red Variant

Profiles of the red variant fermentation are presented in Figure 11, Figure 12 and Figure 13. Comparable profiles for the white variant are presented as Figure 63, Figure 65 and Figure 67 respectively. These white variant profiles were chosen rather than those presented at the beginning of the chapter because the cells were grown under the same conditions as used for the red variant. Both sets of data were collected from fermentations performed in a 20 litre bioreactor. The profiles seem to be somewhere in-between those displayed by the white variant and CA340. For CA340 see Figure 49, Figure 51 and Figure 53.

The red variant did not behave as expected. Under conditions of nitrogen limitation high levels of red pigment production were expected (see Ushio, 2003). However under these conditions no red pigment was formed. When the culture was plated out it was found that some colonies produced the pigment and some did not, none sporulated. This suggests that some form of partial spontaneous reversion may have occurred early in the process. These kind of event occur with a reasonably high frequency in *S. erythraea* (see Seno and Hutchinson, 1986). In all other respects the strain behaved like the red variant with lower growth rate and erythromycin production than the white variant.

The OUR and CER readings for the red variant are lower than for the white variant but higher than for CA340. The stationary phase is also longer than that of the white variant and has a less distinct end (in CA340 the stationary phase is longer and has no distinct end). Organic acid production and trehalose accumulation is lower than those obtained with the white variant but higher than those obtained with CA340. Apart from the trehalose profile the biomass profiles are broadly similar to those seen in the white variant. The red variant has a maximum specific growth rate of 0.067 h^{-1}, compared to 0.09-0.1 h^{-1} for the white variant and 0.046-0.05 h^{-1} for CA340.

5.3.3.2 Comparison of Flux Distributions for White Variant and Red Variant

The data collected from the fermentation was used to perform flux balance analysis of the fermentation. Flux distributions representative of the growth phase and the stationary phase are presented as Figure 68 and Figure 69. The flux distributions follow similar patterns to those seen in the white variant in both phases. The main difference is that the fluxes seem to be smaller in the red variant.

In order to aid comparison the profiles of 5 key fluxes from the red variant and the white variant are plotted as Figure 70 through to Figure 75. The red variant data is compared with that from white variant fermentation 3 (see section 3.3.3). Direct comparison has not been made with strain CA340 as this was grown in a different model and size of bioreactor.

In terms of its flux distributions the red variant also shows an intermediate position between that seen in the white variant and that seen in CA340. Its glucose uptake rate is lower than the white variant's throughout the fermentation allowing a longer stationary phase (see Figure 70). The rate of glucose uptake at a particular glucose concentration is lower in the red variant than in the white variant (see Figure 71). The shape of the profile does however show more similarity with the white variant in the stationary phase than with CA430. This reduced glucose uptake rate gives the red variant a longer stationary phase than the white variant. No erythromycin production is seen in the exponential phase even though the glucose uptake rate and growth rate are lower in this strain than in the white variant. This suggests that the limitation is not severe enough to trigger secondary metabolism.

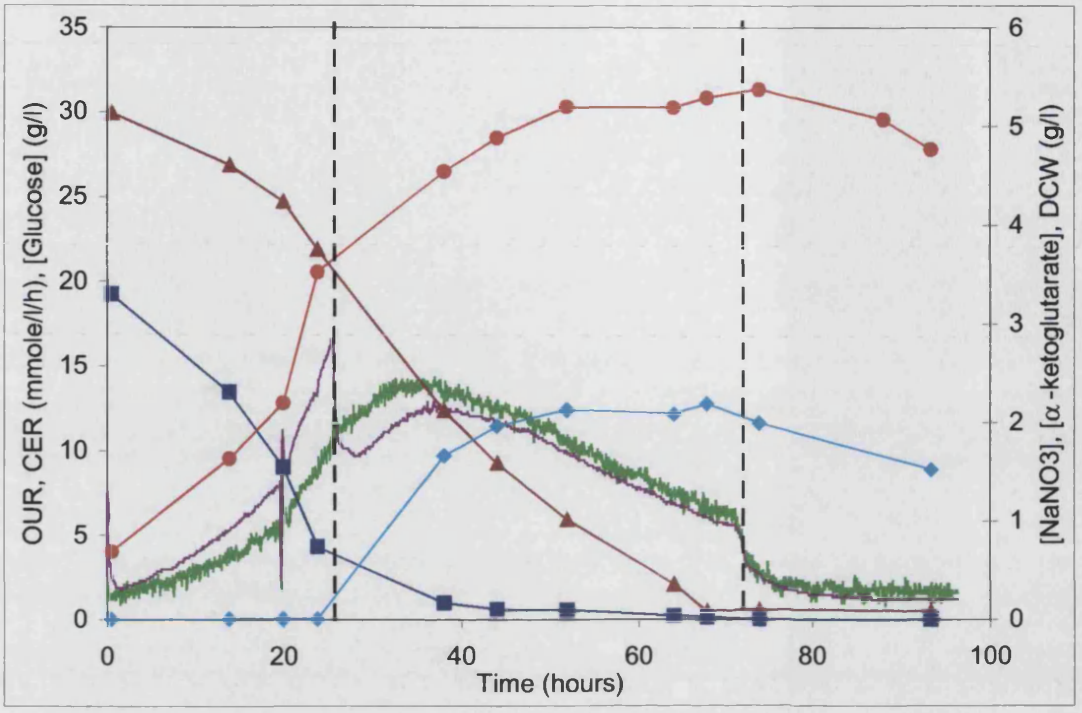


Figure 62: Red variant wild type fermentation 1. OUR (mmole/l/h) —, CER (mmole/l/h) —, Glucose (g/l) ▲, Nitrate (g/l) ■, α-Ketoglutarate mono sodium salt (g/l) ◆, Dry cell weight (g/l) ●. Dashed lines indicate the end of the growth phase and the depletion of glucose from left to right respectively.

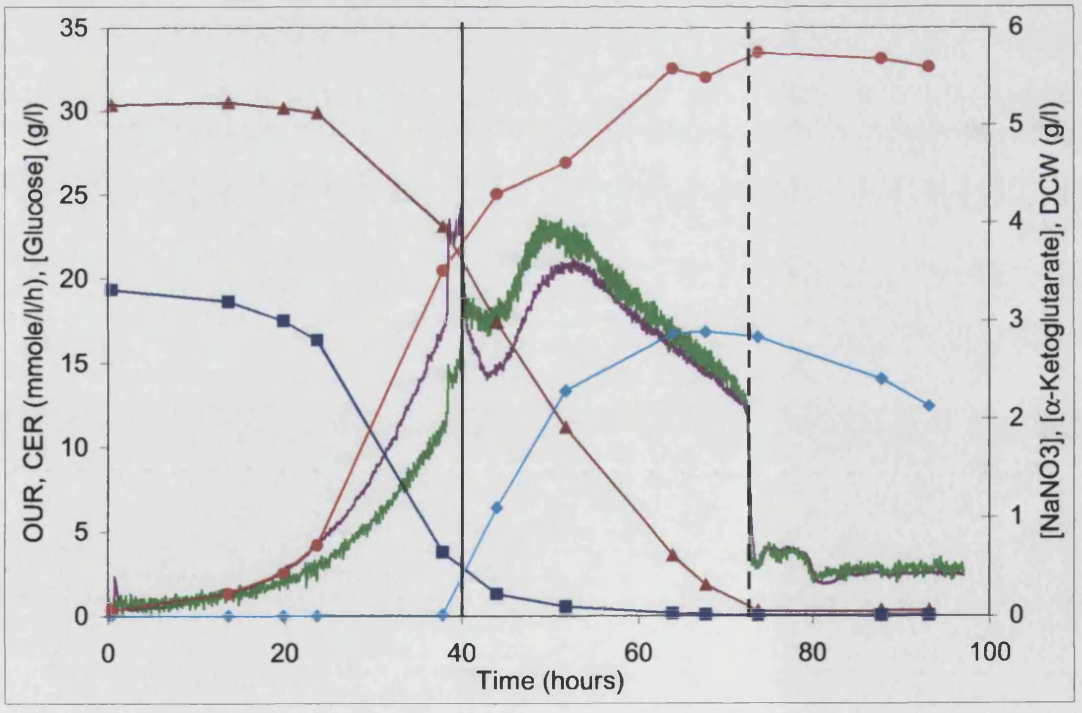


Figure 63: White variant Wild type fermentation 3. OUR (mmole/l/h) —, CER (mmole/l/h) —, Glucose (g/l) ▲, Nitrate (g/l) ■, α-Ketoglutarate (g/l) ◆, Dry cell weight (g/l) ●. Dashed lines indicate the end of the growth phase and the depletion of glucose from left to right respectively.

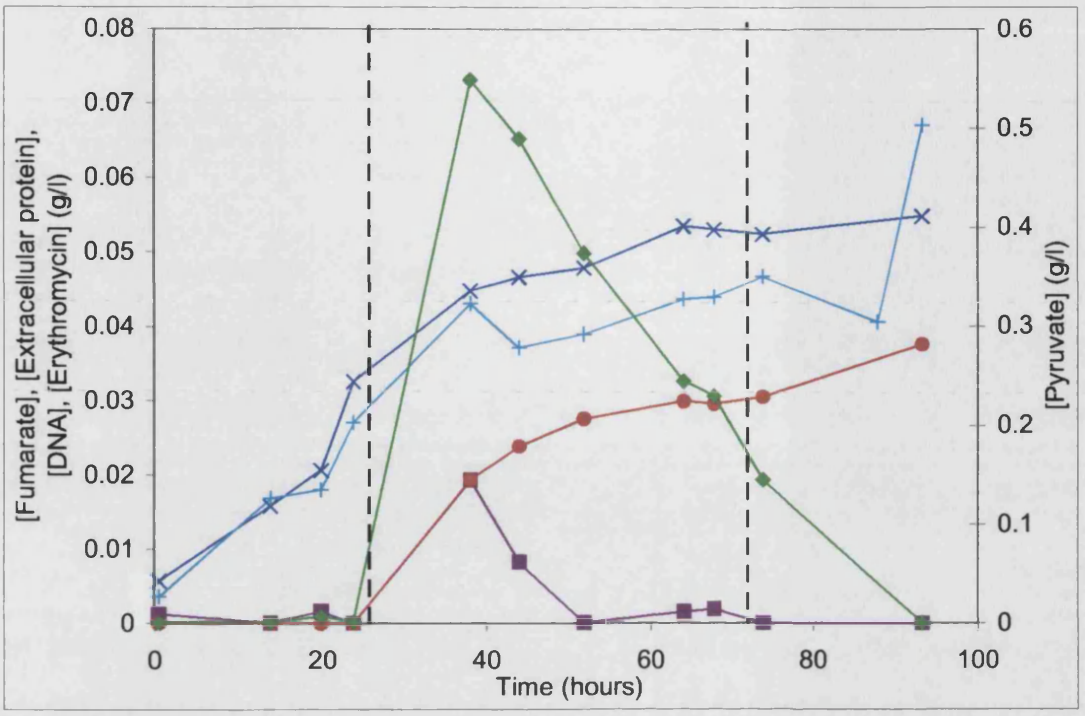


Figure 64: Red variant wild type fermentation 1. Fumarate disodium salt (g/l) ■ , Extra cellular protein (g/l) × , Erythromycin (g/l) ● , DNA (g/l) + , Pyruvate (g/l) ◆. Dashed lines indicate the end of the growth phase and the depletion of glucose from left to right respectively.

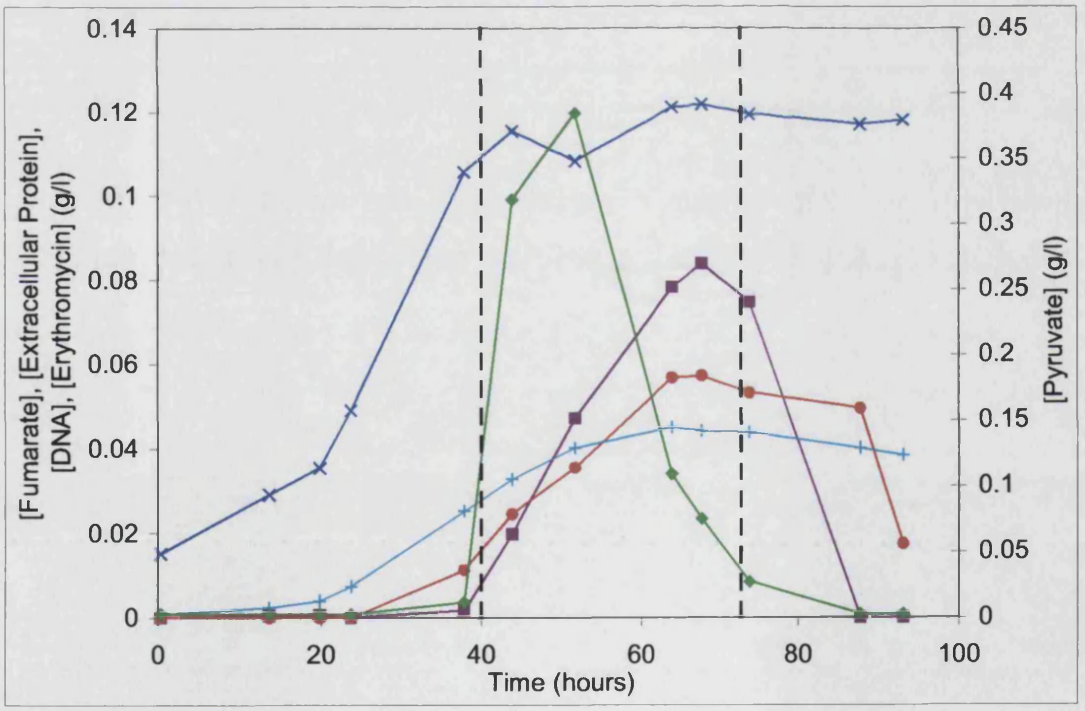


Figure 65: White variant wild type fermentation 3. Fumarate disodium salt (g/l) ■ , Extra cellular protein (g/l) × , Erythromycin (g/l) ● , DNA (g/l) + , Pyruvate (g/l) ◆. Dashed lines indicate the end of the growth phase and the depletion of glucose from left to right respectively.

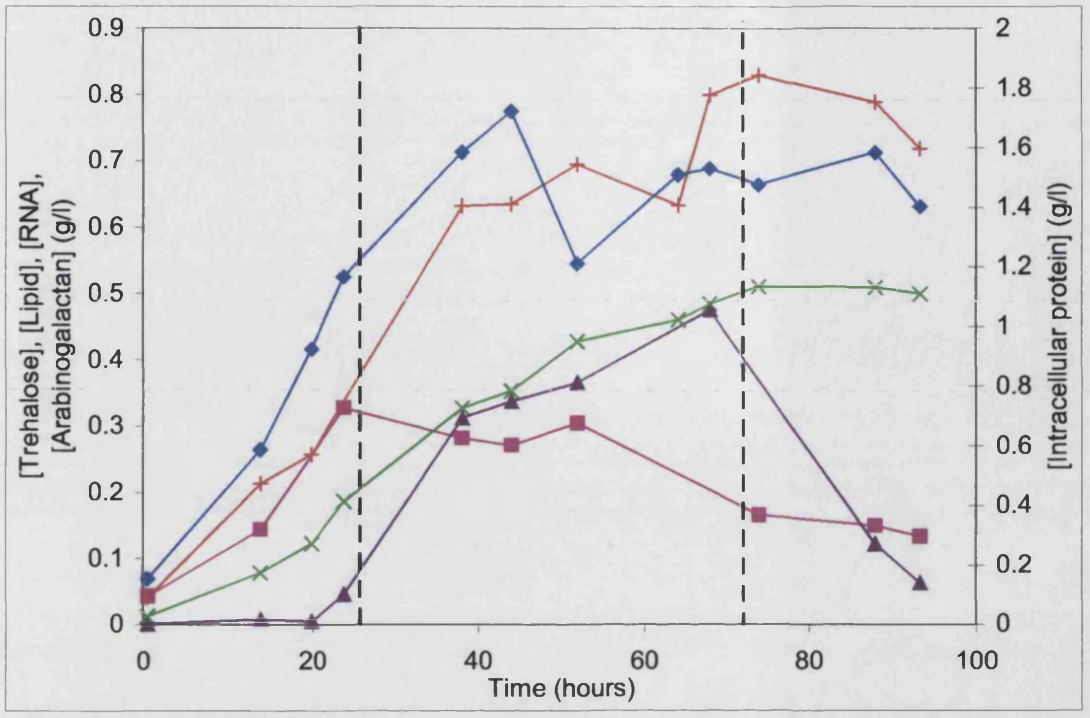


Figure 66: Red variant wild type fermentation 1. RNA (g/l) ■ , Lipid (g/l) × , Trehalose (g/l) ▲ , Arabinogalactan (g/l) + , Intracellular protein (g/l) ◆. Dashed lines indicate the end of the growth phase and the depletion of glucose from left to right respectively.

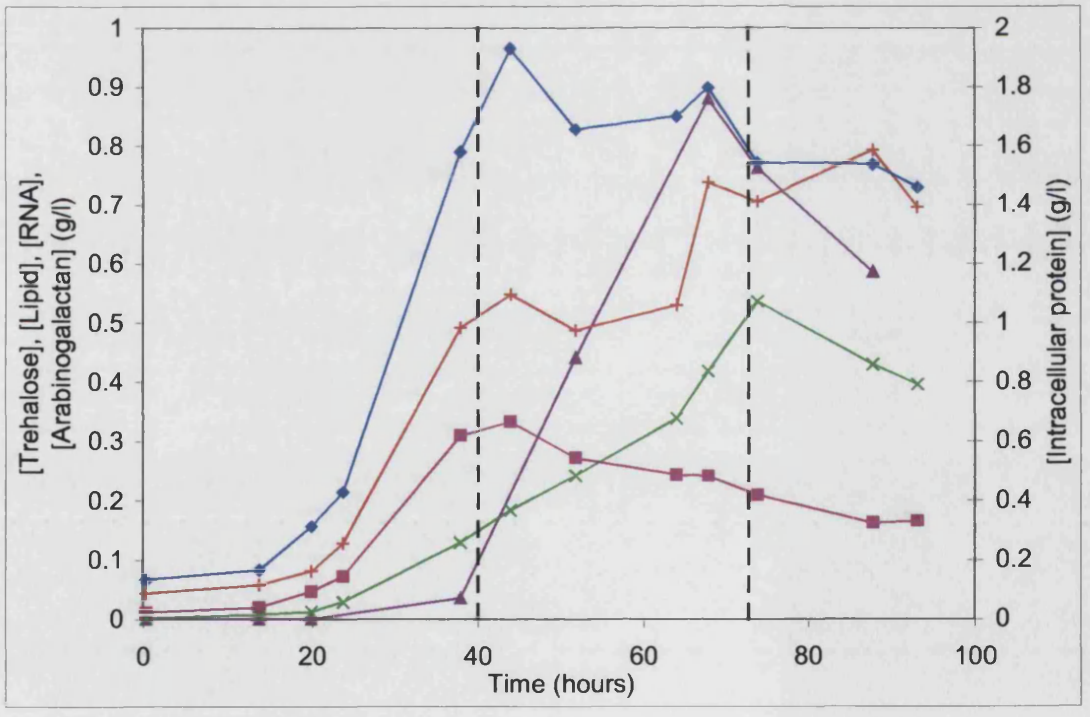


Figure 67: White variant Wild type fermentation 3. RNA (g/l) ■ , Lipid (g/l) × , Trehalose (g/l) ▲ , Arabinogalactan (g/l) + , Intracellular protein (g/l) ◆. Dashed lines indicate the end of the growth phase and the depletion of glucose from left to right respectively.

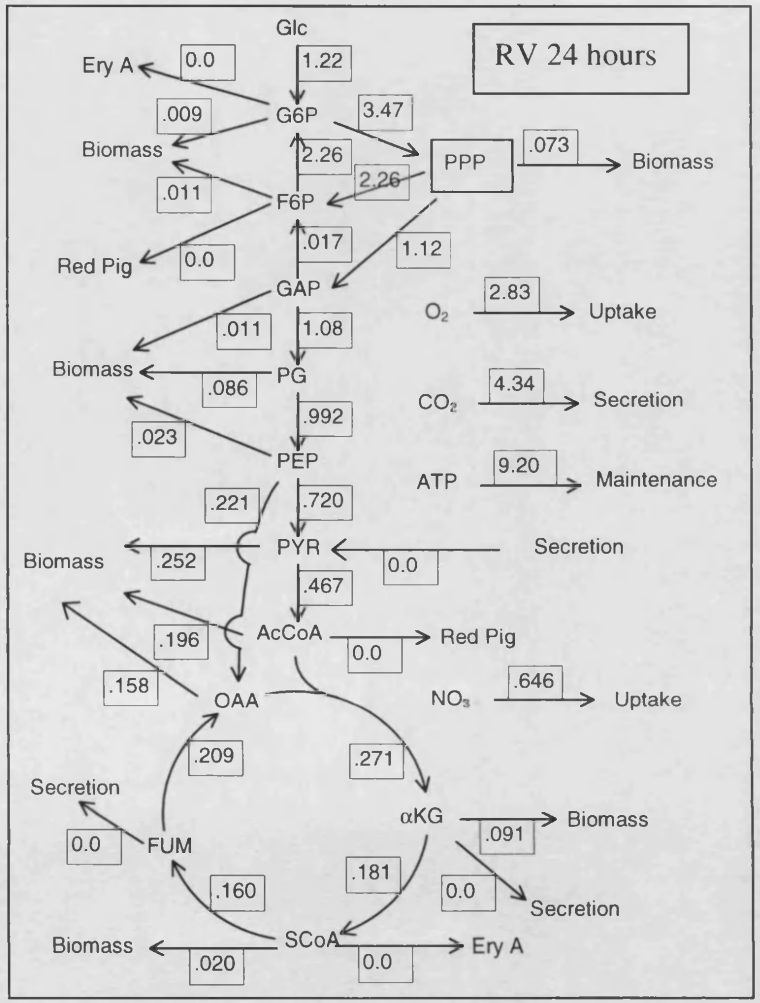


Figure 68: Red variant growth phase flux distribution at 24 hours. Nitrate flux undefined, weighted residual = 1.4. All fluxes are in mmole/gDCW/hour

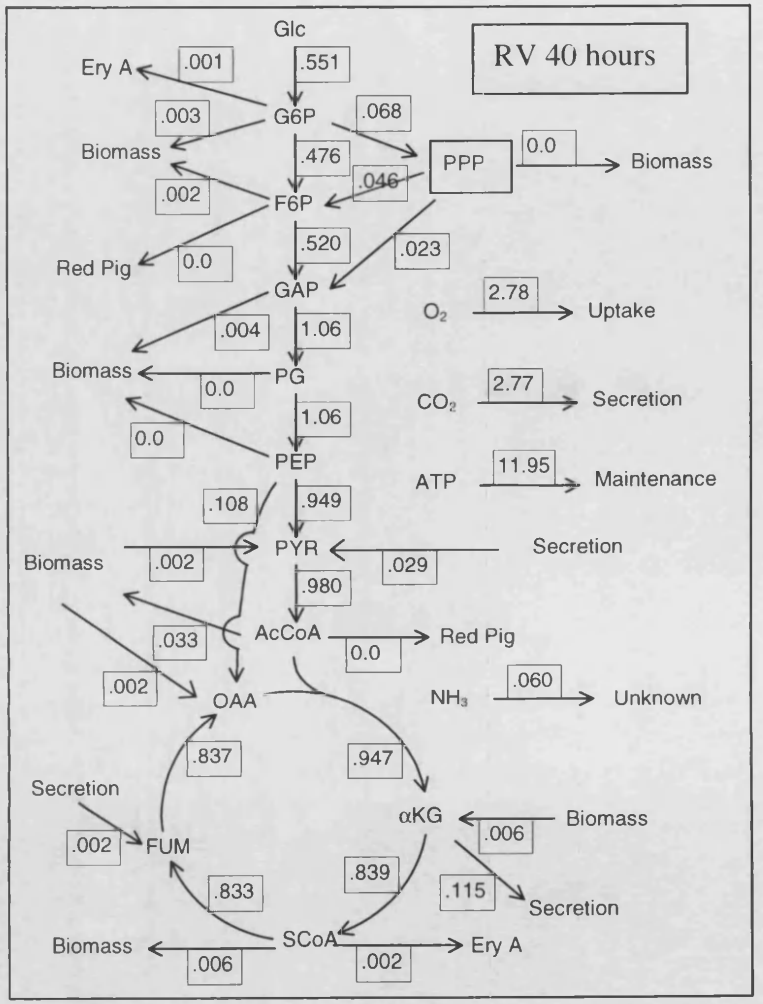


Figure 69: Red variant stationary phase flux distribution at 40 hours. Flux of biomass to ammonia undefined, residual = 2.8. All fluxes are in mmole/gDCW/hour

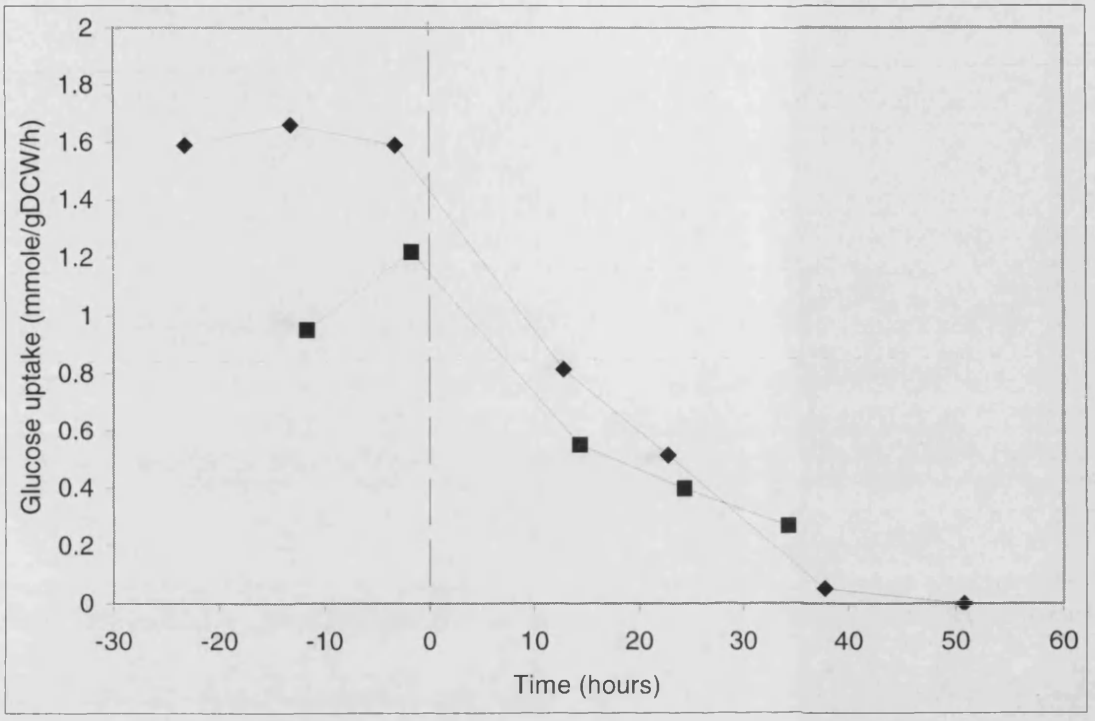


Figure 70: Comparison of glucose uptake rates between the red variant wild type fermentation 1 and the white variant wild type fermentation 3. White variant fermentation ◆, Red variant fermentation ■. The profiles have been aligned by defining the depletion of nitrate as zero hours, this is marked by a dashed line.

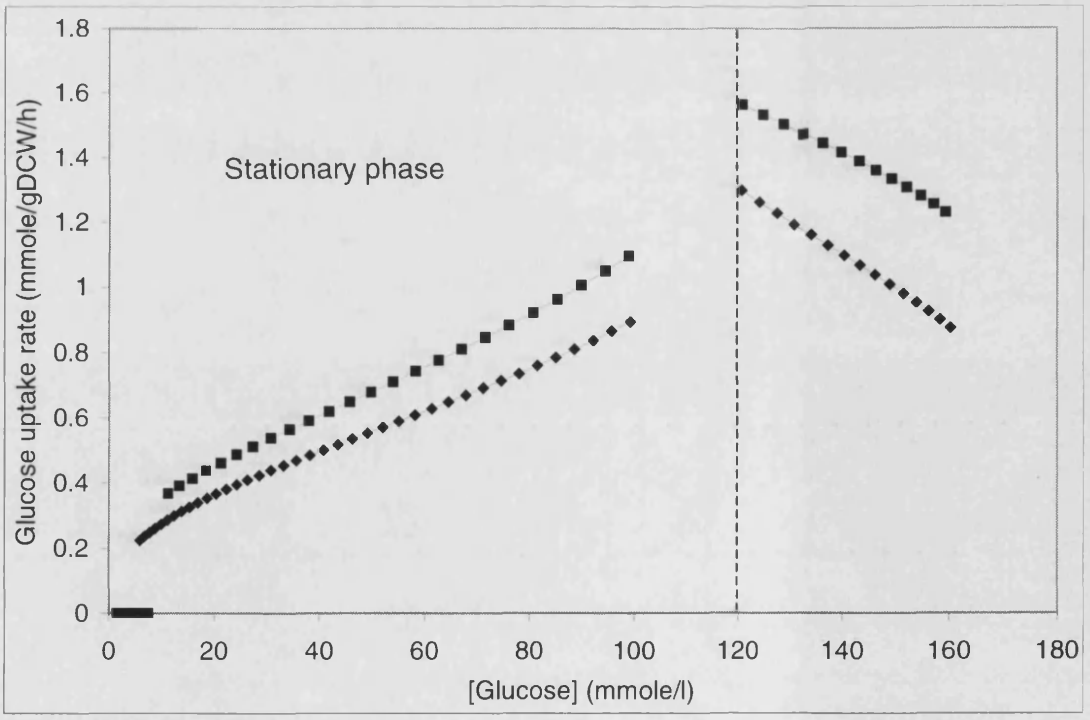


Figure 71: Variation of glucose uptake rate with glucose concentration for white variant wild type fermentation 3 and red variant fermentation 1. White variant ■, Red variant ◆. Dashed line marks the end of growth phase for both strains.

The flux through the PPP is lower during the growth phase for the red variant than for the white variant although by the end of the exponential phase it is approaching that of the white variant (see Figure 72). The same drastic drop in flux through the PPP is seen when the nitrate runs out and the low levels of flux seen in the stationary phase are similar between the two strains.

The flux through phosphoenolpyruvate carboxylase is lower in the growth phase for the red variant but rises towards the level of the white variant by the end of the growth phase (see Figure 73). During the stationary phase the anaplerotic fluxes are similar, however the flux for the white variant descends faster as it runs out of carbon faster. The depletion of carbon and the rate of anaplerotic flux are linked. The anaplerotic flux supplies the precursor of α-ketoglutarate, which uses up the carbon available to the cell. By channelling carbon to α-ketoglutarate more quickly the white variant depletes the glucose faster and has a shorter stationary phase. This causes the steeper decent of the anaplerotic flux in this strain.

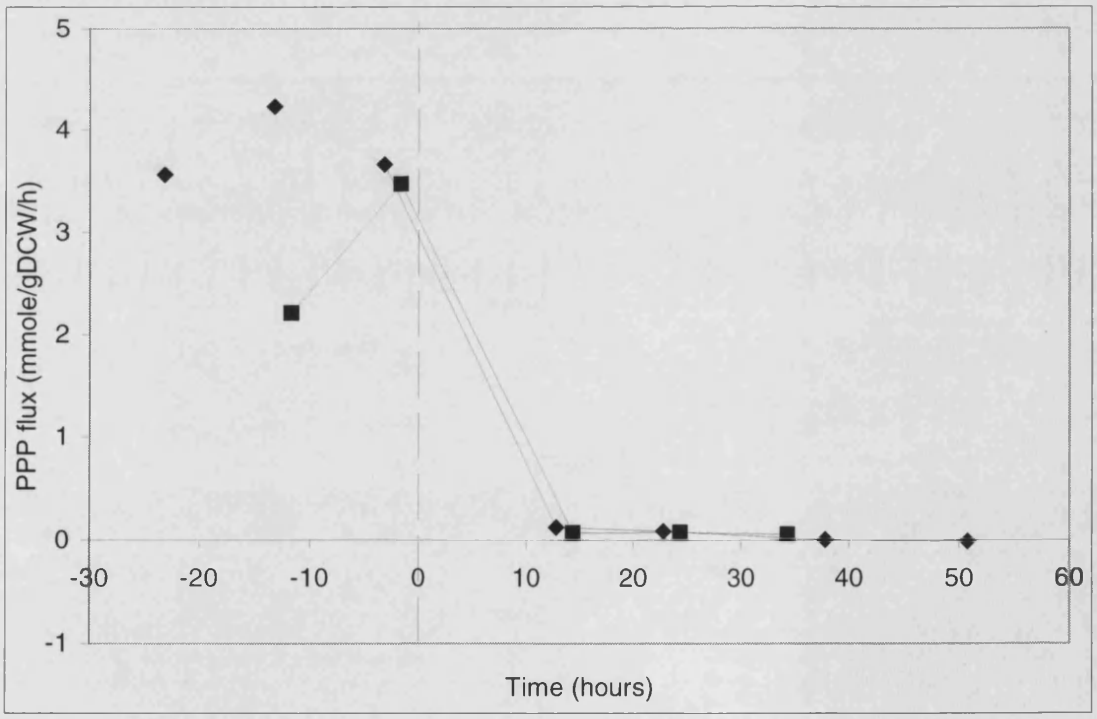


Figure 72: Comparison of calculated PPP flux between the red variant wild type fermentation 1 and the white variant wild type fermentation 3. White variant fermentation ◆, Red variant fermentation ■. The profiles have been aligned by defining the depletion of nitrate as zero hours, this is marked by a dashed line.

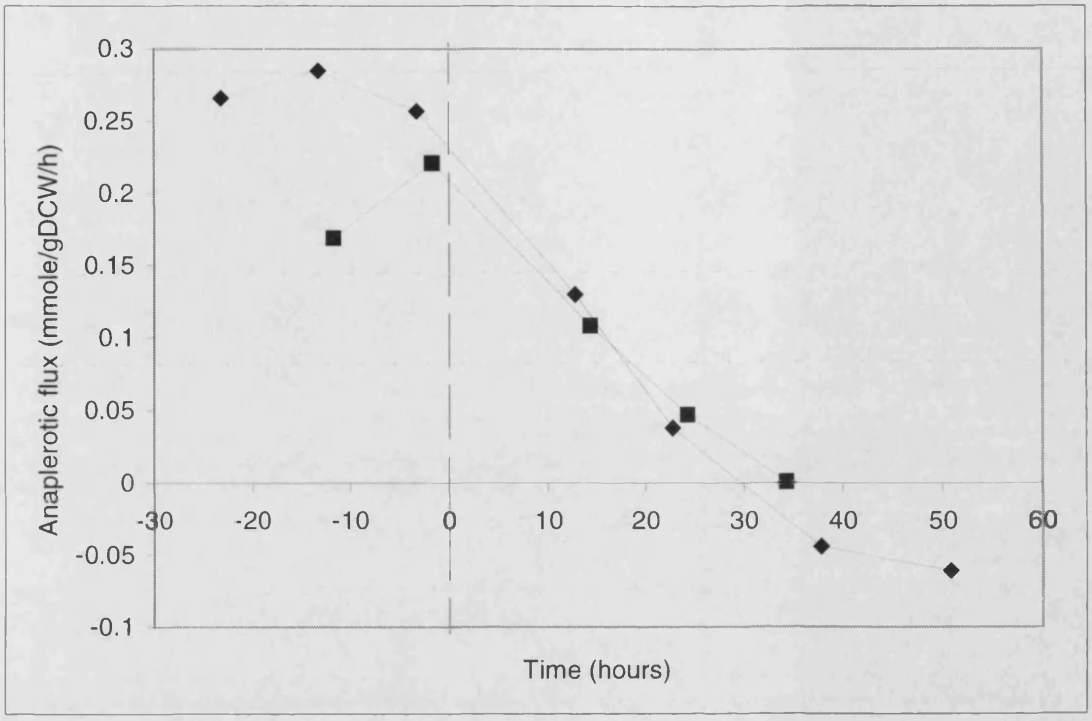


Figure 73: Comparison of calculated anaplerotic flux between the red variant wild type fermentation 1 and the white variant wild type fermentation 3. White variant fermentation ◆, Red variant fermentation ■. The profiles have been aligned by defining the depletion of nitrate as zero hours, this is marked by a dashed line.

The TCA cycle profiles of the red and white variants show a similar pattern. Fluxes are approximately constant in the growth phase, increasing as the stationary phase starts and then steadily declining through the stationary phase as the glucose concentration declines (see Figure 74). The red variant however has much lower TCA cycle activity than the white variant particularly in the exponential phase. This seems to be due to having a lower maintenance energy (see Figure 75). This difference in the maintenance energy requirement is seen through out the fermentation until the white variant runs out of glucose. At that point in the white varient the amount of ATP being used for maintenance purposes falls drastically. In the growth phase the PPP produces large amounts of both NADPH and NADH with the TCA cycle providing the rest of the NADH requirements. However the lower maintenance energy requirements in the red variant mean that the PPP can supply most of the NADH requirements in the growth phase leading to very low TCA cycle flux. When growth halts the PPP activity is greatly reduced and hence also the NADH produced by this route. The TCA cycle activity goes up as the cell enters stationary phase in order to supply the required NADH which the PPP is no longer supplying.

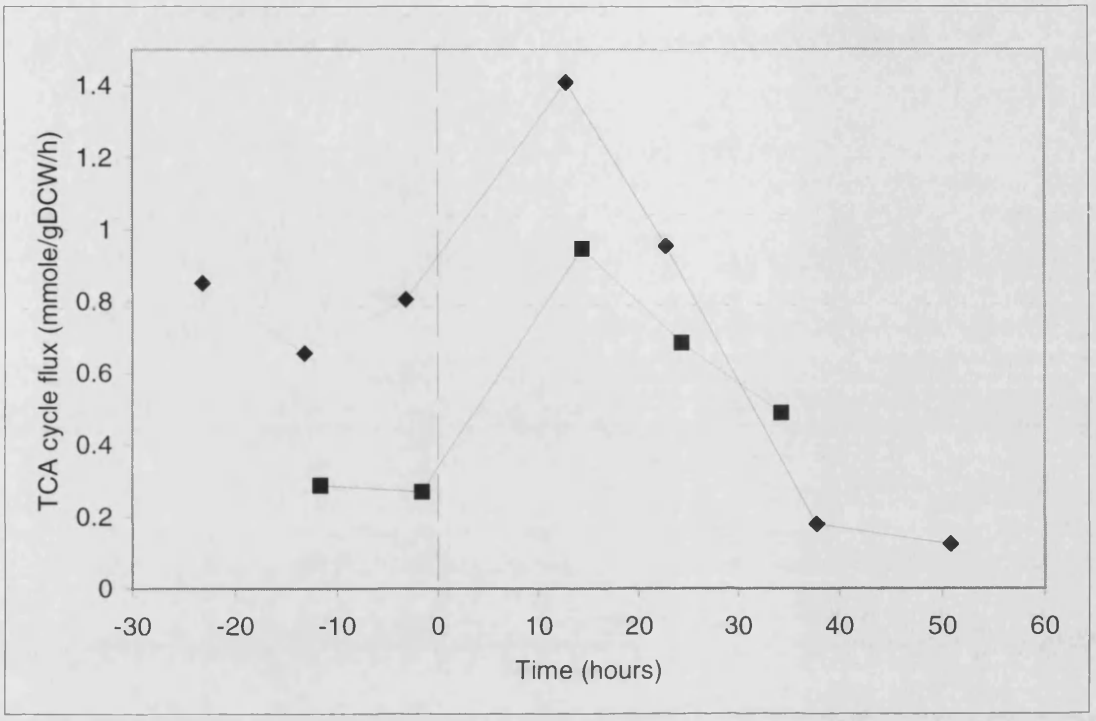


Figure 74: Comparison of calculated TCA cycle flux between the red variant wild type fermentation 1 and the white variant wild type fermentation 3. White variant fermentation ◆, Red variant fermentation ■. The profiles have been aligned by defining the depletion of nitrate as zero hours, this is marked by a dashed line.

In a manner analogous to CA340 many of the differences between the red variant and the white variant can be explained by the difference in the glucose uptake rate. It grows more slowly and seems to waste less carbon in the production of organic acids and in excessively high maintenance utilisation of ATP. The differences are however less extreme than those seen with CA340.

The red variant, although displaying greater prudence in its use of carbon, produces less erythromycin than the white variant. It is not possible to make a tight link between general changes in the metabolic patterns in the cell and production of a specific product. Although the metabolism of the red variant may be more favourable to higher yields of erythromycin than the white variant that does not mean that more erythromycin will be made. These same conditions are also favourable to the formation of a number of other products such as lipid and trehalose. The extra potential can be channelled elsewhere in the cell.

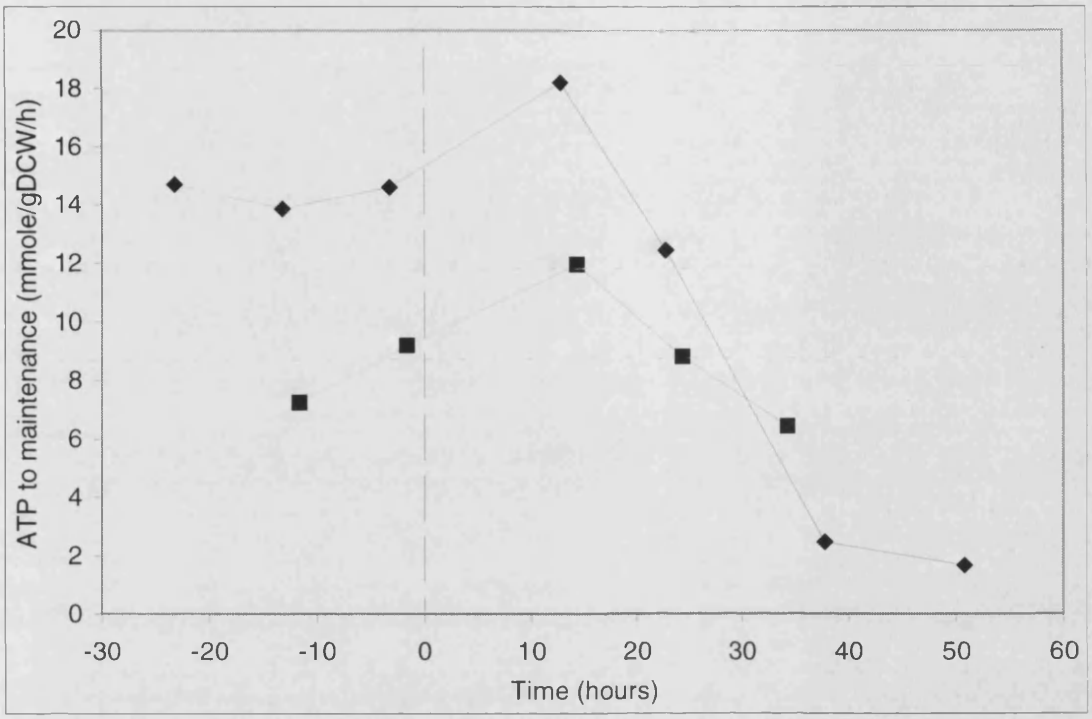


Figure 75: Comparison of calculated maintenance requirements or ATP between the red variant wild type fermentation 1 and the white variant wild type fermentation 3. White variant fermentation ◆, Red variant fermentation ■. The profiles have been aligned by defining the depletion of nitrate as zero hours, this is marked by a dashed line.

The red variant shows a lower growth rate than the white variant but higher growth rate than strain CA340. It was speculated earlier that carbon limitation might be causing this lower growth rate in CA340 and also causing the production of erythromycin during the growth phase. The red variant does not show any synthesis of erythromycin in the growth phase. It may be that there is a threshold level of glucose limitation/growth rate limitation beyond which erythromycin production occurs.

In section 5.2.3 a theory was formulated as to how the properties of strain CA340 might lead to overproduction of erythromycin. This section has shown that although the red variant had some of these properties it did not overproduce erythromycin when compared to the white variant. Introduction of general changes to favour overproduction of a product may not be enough on its own to achieve overproduction. These general properties may have to be linked to specific changes for overproduction of the desired product if this approach is to be successful. It ought to be noted however that in this medium CA340 also produced less erythromycin than the white variant.

5.4 Assessment of Reverse Metabolic Engineering

Reverse metabolic engineering studies like this can help identify strategies for overproduction of secondary metabolites from strains that have been selected for high titres in industry. However the real interest of this is in increasing the titres of new strains of related organisms which have not undergone classical strain improvement programs. This will require the identification of general principles for increasing titres rather than organism specific strategies which cannot be transferred between organisms.

If the strategy devised above of restricting glucose uptake turns out to be effective it is a strategy that is likely to prove general for similar organisms which produce secondary metabolites when exposed to carbon limitation. It also ought to be a reasonably simple modification to make in other organisms as it involves only one step, glucose uptake, and that is a well studied step. This however is not likely to lead to huge levels of secondary metabolite overproduction. It may however lead to modest but important increases in production relatively quickly. The complexity of the metabolism of living organisms however means that even general approaches may not transfer well from one organism to another. It may also be the case that the lowered glucose uptake rate is one of a number of changes which act together to increase erythromycin production in strain CA340. It is however interesting that slowly released sources of carbon are often employed in industrial fermentations for antibiotic production (Stanbury *et al.* 1995).

Observations such as that the carbon taken up is being utilised more efficiently with proportionally less being used to meet the cells energy requirements and more being channelled into biomass and product formation are less easy to apply to other systems. It is not clear in this case what the cause of this effect is. This will make transferring the effect difficult. It could be a consequence of the lower glucose uptake rate, but if it has independent origins these are not revealed by flux analysis and another approach will be required to elucidate them.

There is a limit to how much will be achievable by reverse metabolic engineering for overproduction of secondary metabolites. Classical selection methods increase product formation by manipulating and overcoming the control structures in place in the cell. However they have proved unable to approach theoretical yields of secondary

metabolites. To achieve theoretical yields of product would require a totally new control structure for metabolism. This is unlikely to be arrived at by selection as it is difficult to apply strong selective pressure for secondary metabolite formation. Reverse metabolic engineering is limited to applying the adaptations arrived at by classical strain improvement means or that are present in wild type organisms. So while it is of utility in achieving modest increases routes to achieving theoretical yields of antibiotics will probably not be achieved using this approach. Moreover this method depends on copying improvements from industrial strains which may have been selected into local optima which are incompatible with further improvement. In this case introducing these changes into a wild type strain may be limiting the strain to an overproduction strategy that will make further development difficult.

Flux balance analysis has played a fundamental role in the study of the differences between these three strains. It could be argued however that all the differences found are attributable to differences apparent in the fermentation profiles and flux balance analysis was not needed to identify them. This is true, however, flux balance analysis is a powerful tool which allows differences throughout the whole metabolic network to be visualised simultaneously and directly. This allows a clearer idea of the impact of the observed profile differences to be obtained. Furthermore the information provided by flux analysis is quantitative. This means that the impact of differences between strains is measurable providing a better basis for assessment of changes than rule of thumb deductions from visual inspection of data. This aids in the understanding of the causes and consequences of the differences observed between the strains.

Flux balance analysis is however somewhat limited for analysis of secondary metabolism because the pathways of secondary metabolism generally have much smaller fluxes than those of central metabolism. While fluxes in the larger pathways seem to be quantified reasonably easily, small errors in these large pathways propagate large errors in the smaller pathways. Various techniques have been used in this project to minimise these errors however they are still a problem. So for example with perfect data sets it would be possible to assess the contribution of various different routes to propionyl-CoA and methylmalonyl-CoA in the different strains. However the calculated fluxes in these pathways are so small that they are dominated by small

changes in central metabolism and lipid formation. This makes reliable identification of differences in these pathways between two strains more or less impossible.

5.5 Conclusions

Strain comparison has identified three characteristics of the industrial strain which are likely to contribute to its increased titres of erythromycin. Firstly its slow uptake of glucose and secondly its higher efficiency with energy and carbon and thirdly its greater production of NADPH. One of these characteristics seems readily applicable to other potential polyketide producing strains i.e. slowing glucose uptake. However as the red variant demonstrates the introduction of general conditions favourable to overproduction will not necessarily lead to overproduction without the necessary specific stimuli.

Comparison of the wild type and CA340 has suggested that although CA340 may have improved antibiotic producing capabilities compared to the wild type these may represent a local optimum. This may prevent the strain being readily developed to find a global optimum for erythromycin production.

Reverse metabolic engineering has led to the identification of a strategy for the overproduction of erythromycin. This is based on the differences in the general metabolism of the wild type organism and an industrial overproducer. The strategy identified may be generally applicable to related organisms.

Limitations of reverse metabolic engineering have also been identified for secondary metabolite overproduction. Firstly that industrial overproducer strains whilst giving insight into modest overproduction are unlikely to yield strategies for obtaining theoretical yields. Secondly that flux balance analysis whilst yielding insight into central metabolism has problems with pathways with low fluxes due to error propagation from larger pathways. Thirdly to perform flux balance analysis defined media has to be used, however industrial strains are almost always grown on complex media. The results obtained on defined media may not be representative of what happens on complex media.

6 INVESTIGATION OF IMPACT OF NADH LEVELS ON ORGANIC ACID PRODUCTION

6.1 Introduction

In chapter 4 a comparison of the growth phase and stationary phase flux distributions of *S. erythraea* in batch culture was made. This suggested that stationary phase organic acid production might be due to two effects. Firstly the uncontrolled uptake of glucose. And secondly the limitation of flux through the TCA cycle by oxidative phosphorylation mediated by the relative levels of NAD^+ and NADH. In this chapter the link between oxidative phosphorylation, TCA cycle activity and organic acid secretion will be investigated further using an uncoupler of oxidative phosphorylation.

One of the most striking features of secondary metabolism under nitrogen limitation in *S. erythraea* is the secretion of large amounts of organic acids. These organic acids represent far more carbon than the desired product erythromycin. Production of organic acids has been observed in a number of *streptomyces* species (see section 1.5.3.2.2). It is however often growth associated in these organisms rather than being seen exclusively in the stationary phase as is the case in *S. erythraea.* The details of what causes the overproduction of organic acids by the cell are unclear. If the cause could be ascertained it might be possible to devise strategies to use this carbon more efficiently. In chapter 4 it has been suggested that the rate of glucose uptake is independent of the demand for carbon. If there is a limit on how much carbon the cell can use this might explain the secretion of organic acids in the stationary phase.

α-Ketoglutarate and pyruvate are the main organic acid secreted by *S. erythraea* during the stationary phase and α-ketoglutarate is not reconsumed during antibiotic production. M. Ushio (2003) overexpressed α-ketoglutarate dehydrogenase in *S. erythraea* in order to increase the flux to succinyl CoA, a potential precursor of erythromycin biosynthesis (see Figure 84). She increased the *in vitro* activity ten fold, this implies a ten fold increase in the expression of α-ketoglutarate dehydrogenase. This led to a 50%

decrease in α-ketoglutarate secretion. She had previously calculated the flux split at this point in the TCA cycle during maximum α-ketoglutarate production for the wild type organism. Only 4% was secreted as α-ketoglutarate, the remaining 96% continued in the TCA cycle. A 50% decrease in α-ketoglutarate secretion represents a change in the flux split such that 98% continues in the TCA cycle and 2% is secreted as α-ketoglutarate. This is an α-ketoglutarate dehydrogenase activity increase *in vivo* of approximately 2%. However this has been achieved by a 10 fold (1000%) increase in α-ketoglutarate dehydrogenase expression. A tenfold increase in the concentration of the enzyme has had only a very small impact on the flux through α-ketoglutarate. This is strongly suggestive that the flux through the TCA cycle is rigidly controlled possibly by the state of another part of the network. This could be a direct link via substrate availability or product accumulation or could be mediated via an effector molecule.

The regulation of α-ketoglutarate dehydrogenase and pyruvate dehydrogenase have not been investigated in *S. erythraea*, however they have been investigated in other organisms. Pyruvate dehydrogenase was investigated in *Escherichia coli* by Shen *et al.* (1970). They found that pyruvate dehydrogenase was effected by a number of compounds. Intermediates of glycolysis acted to increase its activity and acetyl-CoA decreased its activity. The main effectors however were the ratio of NAD^+ to NADH and the adenylate energy charge in the cell (see Equation 21). A higher ratio of NAD^+ to NADH increased activity of pyruvate dehydrogenase and higher adenylate energy charge decreased activity of pyruvate dehydrogenase. The adenylate energy charge is usually maintained between 0.8 and 0.9 in *E. coli* (Chapman *et al.* 1971). Energy charges above or below these values are detrimental to the cell because the kinetics of enzymes using these compounds are adversely effected (Walker-Simmons and Atkinson 1977). In the stationary phase *S. erythraea* leaks pyruvate, pyruvate dehydrogenase is therefore unlikely to be limited by low levels of glycolytic intermediates. Rather the limitation probably lies with the ratio of NAD^+ to NADH and the adenylate energy charge on the cell. These two factors are linked by oxidative phosphorylation. It was shown in chapter 4 that in the stationary phase *S. erythraea* wastes energy. This implies that it has excess hence the adenylate energy charge in the cell is likely to be high and the ratio of NAD^+ to NADH is likely to be low.

Equation 21 $$\text{Adenylate energy charge} = \frac{([ATP] + \frac{1}{2}[ADP])}{([AMP] + [ADP] + [ATP])}$$

Weiztman (1972) looked at regulation of α-ketoglutarate dehydrogenase in an *Acinetobacter* species and *E. coli*. He found that in *Acinetobacter* NADH and ATP inhibited α-ketoglutarate dehydrogenase activity but ADP and AMP activated the enzyme. This is rather similar to findings for pyruvate dehydrogenase. In *E. coli* NADH was found to be inhibitory but ADP and AMP had no activating effect. The difference in the regulation mechanisms between *E. coli* and *Acinetobacter* species could be due to differences in requirements under aerobic and anaerobic conditions. *Acinetobacter* are strictly aerobic like *S. erythraea* whilst *E. coli* is a facultative anaerobe. These organisms are only distantly related to *S. erythraea* and the results are from *in vitro* experiments which do not necessarily represent what happens *in vivo*. It seems likely however that similar control mechanisms are employed in the metabolism of *S. erythraea*.

Pyruvate and α-ketoglutarate are both intermediates of central metabolic pathways. Earlier chapters have given indications that glucose uptake may be independent of demand (see section 4.3.4.2). The reduction in the growth rate during stationary phase reduces demand for carbon precursors. It also reduces demand for NADPH and ATP for synthesis of biomass components. The continued high utilisation of glucose and the low demand for biomass components leads to overflow of metabolites. α-Ketoglutarate and pyruvate, the two main secretion products are both precursors for NADH dependent dehydrogenase enzymes in the main glucose catabolic pathways. Both enzymes are multienzyme complexes and two of the units are shared in common between the two enzymes (Voet and Voet 1990). Both complexes contain: a specific dehydrogenase enzyme for their substrate, an adihydrolipoyl transacetylase and a dihydrolipoyl dehydrogenase. The latter two units are identical between the two systems. NAD^+ and NADH interact with the third component which is conserved between the two complexes. Hence there is a good probability that there will be similarities in the regulation of these two enzymes especially with respect to the effect of the ratio of NAD^+ to NADH.

Malate is also a precursor for a dehydrogenase enzyme in the TCA cycle, however no malate secretion is seen. Malate dehydrogenase is not a three subunit multienzyme complex (Voet and Voet 1990). In *Corynebacterium glutamicum* a species closely related to *S. erythraea* two forms exist simultaneously, one which is NADH dependent and one which transfers electrons directly to quinones in the electron transport chain (Molenaar *et al.* 2000). The different type of dehydrogenase and the possibility for a different electron acceptor may explain why malate is not also secreted from the cell. There is however another possible explanation. Fumarate is secreted by the cell during the stationary phase and it is possible that fumarate is secreted in place of malate. The enzyme fumarase which converts fumarate to malate operates close to equilibrium under physiological conditions (Rose *et al.* 1992). It may be that malate accumulates but is not secreted and this causes fumarate to accumulate and be secreted.

Based on the considerations outlined above an hypothesis was formed that the build up of organic acids might be due to reduced activity of dehydrogenase enzymes in the TCA cycle relative to glucose uptake. Glucose uptake has been shown in chapters 4 and 5 to be unregulated in the wild type organism. This reduced activity in the TCA cycle might be due to feedback from the adenylate energy charge of the cell (see Equation 21). As the cell moves from growth to stationary phase there is less demand for ATP for synthesis of biomass hence the adenylate energy charge of the cell should rise. This will lead to reduced oxidative phosphorylation as ADP becomes scarce. This in turn would lead to a build up in NADH levels and lowering of NAD^+ levels. The lowered ratio of NAD^+ to NADH and high adenylate energy charge in the cell would limit the activity of the dehydrogenase enzymes causing build-up and overflow of their precursors. The kind of phenomenon outlined here has been reported before in a *streptomyces* species. Hockenhull *et al.* (1954) inhibited α-ketoglutarate and pyruvate dehydrogenase in *Streptomyces griseus* using the chemical inhibitor arsenite. Inhibition of these enzymes led to the overflow of α-ketoglutarate and pyruvate although in that case it was mainly pyruvate. This shows that limitation of the activity of these enzymes is a plausible explanation for the secretion of organic acids in *S. erythraea*.

The main producer of NAD^+ from NADH in the cell is oxidative phosphorylation. This is also the main producer of ATP from ADP and AMP. The oxidation of NADH by molecular oxygen pumps hydrogen ions out of the cell via the electron transport chain.

The membrane potential produced is then used to drive ATP formation from ADP and inorganic phosphate. The TCA cycle as the main producer of NADH in the stationary phase and oxidative phosphorylation as the main consumer of NADH in the stationary phase are therefore tightly linked. The rate of flux through the TCA cycle should be limited by the rate of oxidative phosphorylation. We have already seen in chapter 4 that this rate may be rather inflated by the wastage of energy in the stationary phase. If this level of wastage is limited then the level of oxidative phosphorylation will be limited by the need to maintain a viable ATP charge in the cell. The TCA cycle flux will then be limited by the flux through oxidative phosphorylation. Any flux through glycolysis in excess of the flux round the TCA cycle will have to be secreted.

A number of published reports seem to support this hypothesis. For example Avignone Rossa *et al.* (2002) found that when *Streptomyces lividans* was fed gluconate instead of glucose production of pyruvate and α-ketoglutarate increased significantly. Feeding gluconate forces all carbon entering the cell to be processed through the pentose phosphate pathway, however glucose-6-phosphate dehydrogenase is bypassed. Glucose-6-phosphate dehydrogenase is the PPP enzyme responsible for the regeneration of NADPH. In *S. lividans* as in *S. erythraea* 6-phosphogluconate dehydrogenase is NAD^+ dependant. In order to produce enough NADPH for the cell to grow fructose-6-phosphate from the oxidative branch of glycolysis had to be directed to glucose-6-phosphate to renter the PPP in order to generate NADPH by glucose-6-phosphate dehydrogenase. This mean that far more NADH was being generated in the PPP when grown on gluconate than when grown on glucose. This will have reduced the requirement for TCA cycle activity to supply NADH and led to greatly increased secretion of pyruvate and α-ketoglutarate.

Another work also lends support to this hypothesis. Surowitz and Pfister (1985) looked at pyruvate secretion by *Streptomyces alboniger* when grown on glucose and found pyruvate was secreted as the biomass grew. They measured the levels of activity of several enzymes in glycolysis and the TCA cycle and found that when grown on glucose the activities of the glycolysis enzymes phosphofructokinase and pyruvate kinase were around five times higher than the TCA cycle enzymes pyruvate dehydrogenase and citrate synthase. They ascribed the secretion of pyruvate to this imbalance of the enzymes of glycolysis and the TCA cycle. This imbalance was only

seen with glucose as carbon source, dextrin did not have the same effect. Likewise pyruvate secretion was only seen with glucose. The situation described sounds rather similar to that seen in this work with *S. erythraea* where glucose uptake appears to be unregulated by the demand from the TCA cycle.

A strategy for testing this hypothesis was devised. Certain molecules can uncouple oxidative phosphorylation. These uncouplers generally dissipate the membrane potential of the cell, which is built-up by reoxidation of NADH to NAD^+ (Voet and Voet 1990). The cell's membrane potential is used to synthesise ATP from ADP. Uncoupling this process reduces the number of ATPs synthesised per NADH oxidised. By applying uncouplers the demand for NADH in the cell can be increased independently of the demand for ATP by changing the P:O ratio. This will raise the NAD^+ to NADH ratio and may also lower the adenylate energy charge as the cell may be unable to maintain it. The increased demand for NADH should lead to an increased flux through the dehydrogenase enzymes in the cell. This will reduce the build up and secretion of the precursors used by these enzymes. At the same time the demand for oxygen for oxidative phosphorylation should increase. Likewise the production of CO_2 from the TCA cycle should also increase. The ratio of consumption of O_2 and production of CO_2 should move closer to the theoretical value for total combustion of glucose i.e. 1. We have found only one example of the use of an uncoupler to study the metabolism of a *streptomyces* species. Hockenhull *et al.* (1954) added an uncoupler to a culture of *Streptomyces griseus*, they found a considerable increase in respiration. They did not however measure the effect on organic acid production. Azide, an inhibitor of the electron transport chain, was found to have no effect on the respiration of the organism.

In this section an attempt will be made to demonstrate that in the stationary phase the activities of dehydrogenase enzymes in the TCA cycle of *S. erythraea* are being limited by the rate of oxidative phosphorylation causing secretion of organic acids. Addition of an uncoupler to stationary phase cells of *S. erythraea* is expected to lead to a reduction in the rate of organic acid production and an increase in CO_2 production and O_2 consumption.

6.2 Materials and Methods

S. erythraea white variant wild type was used in these experiments. Growth was performed under nitrate limitation. During the stationary phase an addition of uncoupler was made. Glucose, nitrate, α-ketoglutarate, pyruvate and in some cases OUR, CER and fumarate were recorded. Methods for growth and measurements taken were performed as recorded in chapter 2. Procedures used only in this section are recorded below. The New Brunswick bioreactors were used in this work.

6.2.1 Uncoupler Addition

The uncoupler, carbonyl cyanide-*m*-chlorophenylhydrazone (CCCP) (Sigma biochemical grade), was added to shake flasks or fermenters to achieve a concentration of 10 μM. Due to the low solubility of CCCP in water it was first dissolved in 1 ml Dimethylsulphoxide (DMSO) (Merck AnalR) and then added by sterile filtration. To the controls an equal volume of DMSO was added. The DMSO caused no detectable change in the behaviour of the cells.

Timing of the addition was based on the online OUR and CER profiles for the bioreactors or on data generated by HPLC analysis of the sample for the shake flasks. This gave a slower response time for the shake flasks.

CCCP was selected as the uncoupler based on Silva *et al.* (2001) who used it with *M. tuberculosis* which has a similar cell wall to *S. erythraea*. They stated that they used a 5 mM concentration. This was however found to be in vast excess of the saturation point in water and immediately stopped all metabolic activity. Further examination of the literature revealed that 4 μM was used by Aínsa *et al.* with *M. tuberculosis* and concentrations in this order of magnitude were found to be suitable for this work suggesting a typographical error in Silva *et al.* (2001).

6.2.2 Measurement of Intracellular Metabolites

A rapid sampling and extraction method was used to measure intracellular levels of α-ketoglutarate based on the method of Gonzalez *et al.* (1997). Samples of 1ml of broth were rapidly taken and chilled by addition to 5ml methanol in a 13ml Röhre tube (Sarstedt) precooled to below –30°C in a dry ice ethanol bath. Samples were stored below –30°C until centrifugation for 5 minutes at 3500 RPM, –10°C (Beckman GS6R). The pellet was washed once with cold methanol and frozen at –20°C until extraction. The pellet was extracted in 5 ml of boiling ethanol for 3 minutes. From this 2 ml was dried in a 2 ml Eppendorf tube under vacuum. The contents were then made up in 0.3 ml of ultra-pure water and analysed for organic acids and glucose by HPLC (see section 2.6.15.2).

6.3 Results and Discussion

6.3.1 Preliminary Shake Flask Experiments to Establish Uncoupler Concentration

Preliminary experiments were performed in shake flasks to get an initial indication of whether the hypothesis was correct and to identify the best timing and concentration of uncoupler for addition. In the first experiment the α-ketoglutarate secretion did show a transitory cessation around the time of uncoupler addition. The addition was too late for studying the pyruvate profile as pyruvate concentration had almost peaked and secretion rates were already low when uncoupler was added. The experiment also showed that addition of the uncoupler has little impact on glucose uptake. This is interesting because in *S. coelicolor* the only glucose uptake system found to date is a proton symport (Bertram *et al.* 2004) which uses the membrane potential of the cell to drive glucose uptake. The disruption of the membrane potential by an uncoupler would be expected to adversely effect glucose uptake. It may be that *S. erythraea* uses a different glucose uptake system than *S. coelicolor.* The results from this experiment are not presented because they are similar to those seen in the next experiment. The experiment

confirmed that addition of CCCP had an impact on organic acid production therefore further experiments followed. In the first experiment 2.5µM CCCP was used. In order to increase the length of time the effect lasted 5µM was used in the next experiment. An addition time closer to the start of stationary phase was also employed so that the impact on the pyruvate profile could be better assessed.

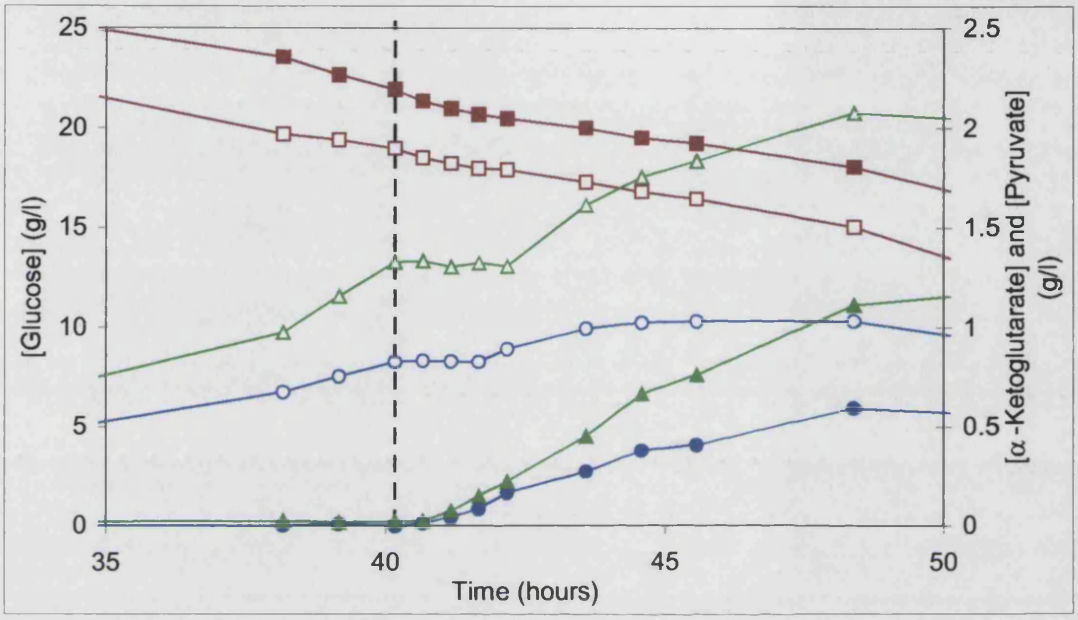


Figure 76. Fermentation profiles for the shake flask uncoupler experiment 2 using 5µM CCCP. The graph shows the time immediately before and after addition of the uncoupler. glucose experiment □, glucose control ■, α-ketoglutarate experiment △, α-ketoglutarate control ▲, pyruvate experiment ○, pyruvate control ●, the dashed line marks the addition of the uncoupler.

Data from the second shake flask experiment is presented in Figure 76. The experiment confirmed the findings of the first experiment with respect to glucose uptake and α-ketoglutarate secretion. This time however the pyruvate secretion also displayed temporary cessation upon addition of the uncoupler. The extra uncoupler added did not seem to increase the length of the pause in α-ketoglutarate secretion. The transitory nature of the phenomenon seems to suggest the cells are able to rapidly adapt to the presence of the uncoupler. The concentration was increased again for the next experiment in order to try to increase the length of the effect.

6.3.2 Effect of Uncoupler on Stationary Phase Cells of *S. erythraea* in Batch Culture

In order to supplement the metabolite profiles found in the shake flask experiments with data regarding the respiration of the cells upon the addition of CCCP the experiments were repeated in fermenters. Four fermentations were run simultaneously. In two CCCP dissolved in DMSO was added to 10μM during the stationary phase and in two only DMSO was added. All fermentations grew although the third reactor showed a long lag phase and low pyruvate secretion. The graphs in this section have been presented such that ten hours is shown starting shortly before the switch from growth phase to stationary phase. This is done in order to allow closer analysis of the key time around the addition of the uncoupler.

6.3.2.1 Description of Results

Profiles from these experiments are presented in Figure 77 to Figure 80, only the portion of the fermentation around the addition of the uncoupler is shown. Figure 77 and Figure 79 are controls where only DMSO was added, Figure 78 and Figure 80 had CCCP added. Figure 77 is a typical fermentation profile for *S. erythraea* with a sharp drop in CER as the nitrogen source becomes depleted. The CER then slowly increases again over the next ten hours. This is the period when organic acid production is highest and as the CER starts to drop again so does the extracellular concentration of pyruvate. This vessel grew slightly faster than the others making the addition time slightly later in the profile. The profile shows no impact upon addition of DMSO. Figure 78 shows a typical profile up until the addition of uncoupler. The addition of the uncoupler impacts the organic acids profile in the same way as seen in the shake flask experiments. The duration of the cessation of production of organic acids is however shorter. The glucose profile shows no impact upon addition of uncoupler as in the shake flask experiments. Both the OUR and the CER show sharp changes upon the addition of the uncoupler. Both drop sharply initially and then recover over the next 30 minutes, over shooting their original level before settling down. The amount of O_2 represented by the sharp drop in the OUR is larger than the amount of CO_2 represented in the sharp drop in the CER. However the amounts of O_2 and CO_2 represented by the overshoot are of similar size.

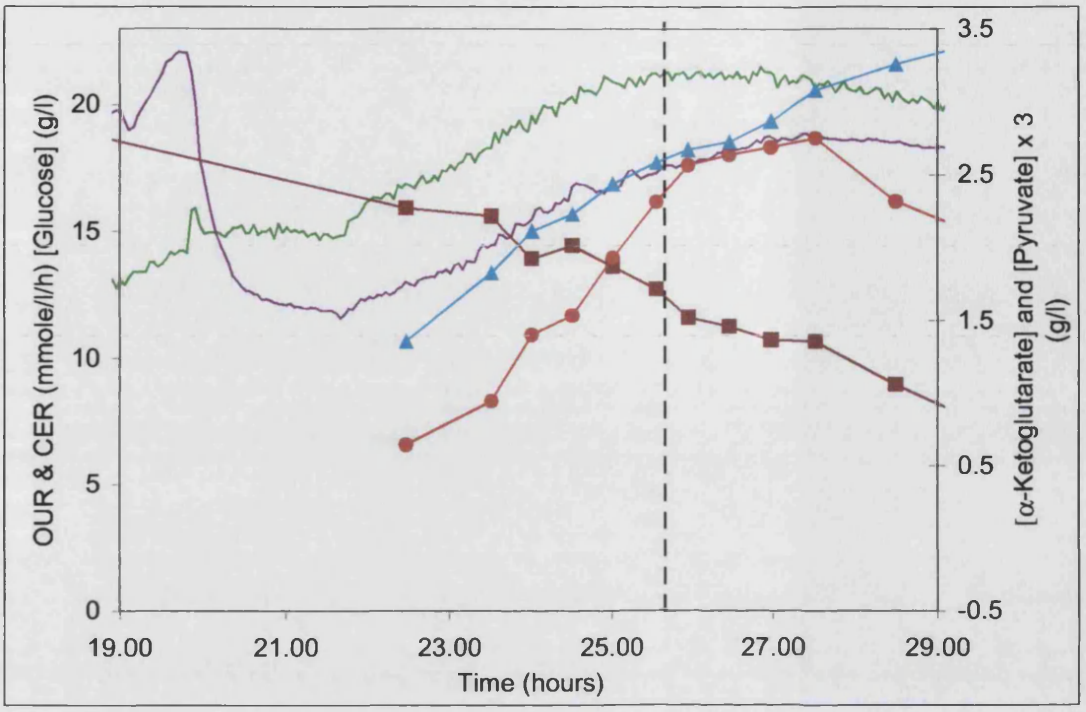


Figure 77. Uncoupler experiment control fermentation 1. This graph shows the profiles of interest around the addition time for DMSO. Glucose ■, α-ketoglutarate ▲, pyruvate ●, OUR —, CER —. The dashed line marks the addition of the DMSO.

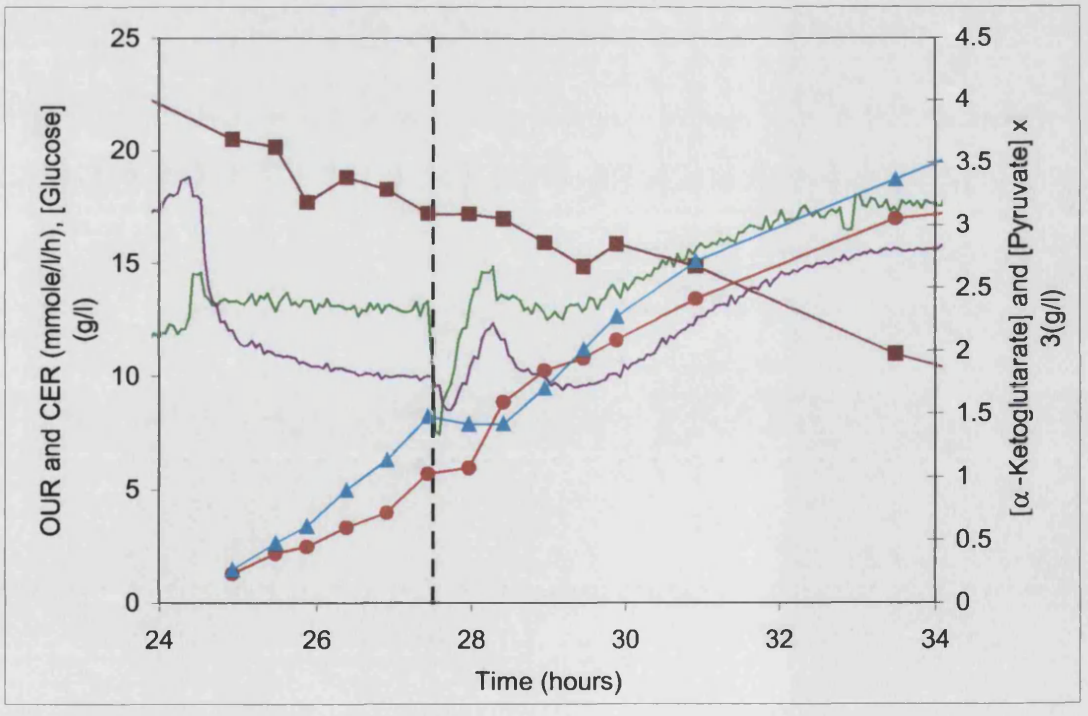


Figure 78. Uncoupler experiment, experimental fermentation 1. This graph shows the profiles of interest around the addition time for the uncoupler. Glucose ■, α-ketoglutarate ▲, pyruvate ●, OUR —, CER —. The dashed line marks the addition of the uncoupler.

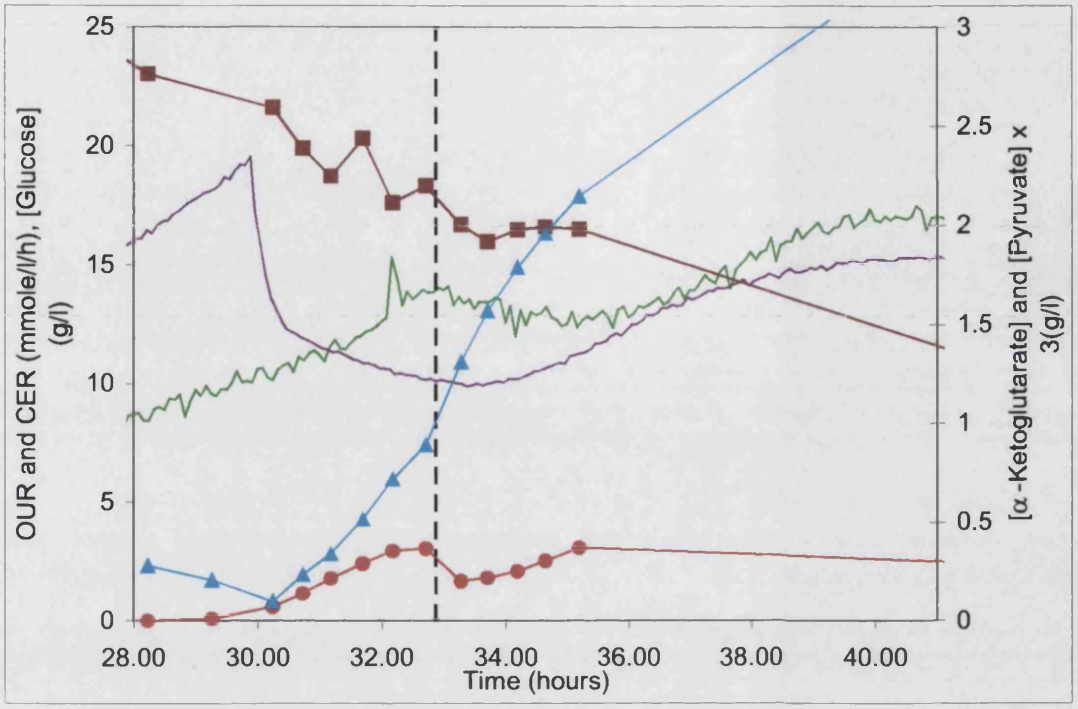


Figure 79. Uncoupler experiment control fermentation 2. This graph shows the profiles of interest around the addition time for the DMSO. Glucose ■, α-ketoglutarate ▲, pyruvate ●, OUR —, CER —. The dashed line marks the addition of the DMSO.

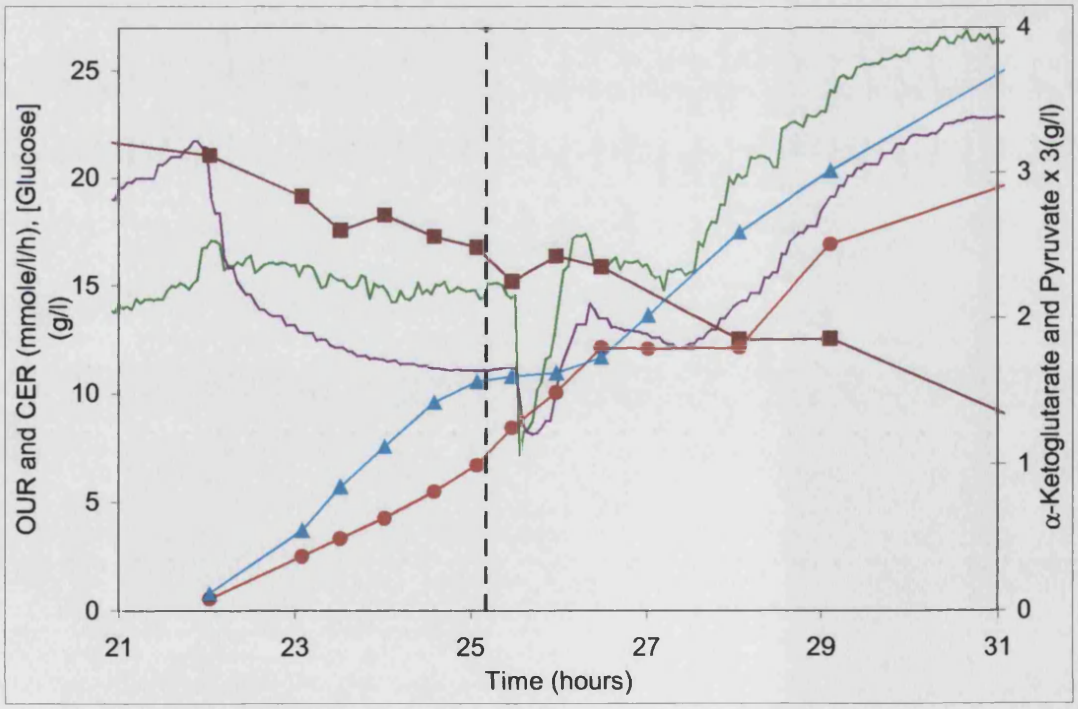


Figure 80. Uncoupler experiment, experimental 2. This graph shows the profiles of interest around the addition time for the uncoupler. Glucose ■, α-ketoglutarate ▲, pyruvate ●, OUR —, CER —. The dashed line marks the addition of the uncoupler.

The responses are similar in Figure 79 and Figure 80. In Figure 80 the discrepancy between the gas profiles and the addition of uncoupler is due to residence time of gas in the pipes between the reactor and the detector. Where known this residence time has been accounted for in the gas profile. In this case it was unknown. The pyruvate profile seen in Figure 80 is different from those seen in shake flasks and in Figure 78. There is still a cessation from pyruvate production but it lags behind the addition of the uncoupler and the α-ketoglutarate pause.

6.3.2.2 Analysis of Results

The organic acid profiles from the shake flasks and the first uncoupler fermentation support the hypothesis that organic acid production is caused by limited availability of NAD^+. However the pyruvate profile from the second uncoupler fermentation does not fit this hypothesis. In this fermentation pyruvate synthesis does not halt on addition of the uncoupler but halts some time later. The gas profiles confuse the issue further. Both the OUR and the CER were expected to go up on addition of the uncoupler. Immediately after the addition of the uncoupler the hypothesis does not hold true. Both actually initially go down although the RQ does approach 1 as predicted by the hypothesis. However later results are more in agreement with the hypothesis. After 20-30 minutes the OUR and CER both increase beyond the pre-uncoupler levels with roughly equal amounts of excess. The secretion of α-ketoglutarate however remains halted suggesting that carbon is being channelled to respiratory activity rather than α-ketoglutarate formation. This is in agreement with the hypothesis that OUR and CER will rise while organic acid synthesis falls, the RQ however is not 1.

To the best of my knowledge no comparable work has been published with which to relate these results. Uncoupler experiments are usually performed on a small scale and measurements such as OUR and CER are not normally taken. Ito *et al.* (1983) measured uptake of oxygen in *E. coli.* Cells were removed from their culturing conditions and resuspended in media containing uncoupler, oxygen levels in solution were measured with a Clark type oxygen electrode. They found on addition of 4μM CCCP the OUR increased almost two fold. Additions of 40μM prevented all respiration. Hockenhull *et al.* (1954) exposed cells of *Streptomyces griseus* to the uncoupler

dinitrophenol. The cells had been removed from the growth medium and starved for 24 hours in phosphate buffer. Addition of the uncoupler was found to increase respiration considerably. Neither work considered the impact on organic acid formation. The results found here are rather different from those found by Hockenhull *et al.* (1954). The results suggest that addition of the uncoupler initially represses respiration. Respiratory activity recovers from this over about 20 minutes and then increases beyond pre-uncoupled levels for a time before the cells recover from the addition of the uncoupler.

In order to confirm these rather unexpected results and to try to establish which of the two pyruvate profiles seen was correct the experiment was repeated.

6.3.3 Confirmation of Effect of Uncoupler on Stationary Phase Cells

6.3.3.1 Analysis of Fermentation Profiles

In this experiment two fermenters were run an experimental and a control. The fermentation profiles are presented in Figure 81 and Figure 82. The profiles agree closely with the previous experiment. Once again the OUR and CER decrease rapidly upon addition of the uncoupler. As before the results from the first few minutes do not fit with the hypothesis. From 20-40 minutes however both profiles overshoot the pre-uncoupler levels with roughly equal excess. This does supports hypothesis concerning organic acid synthesis in the later part of the experiment. The pyruvate profile agrees with the second uncoupler fermentation in that it displays a belated pause in pyruvate production.

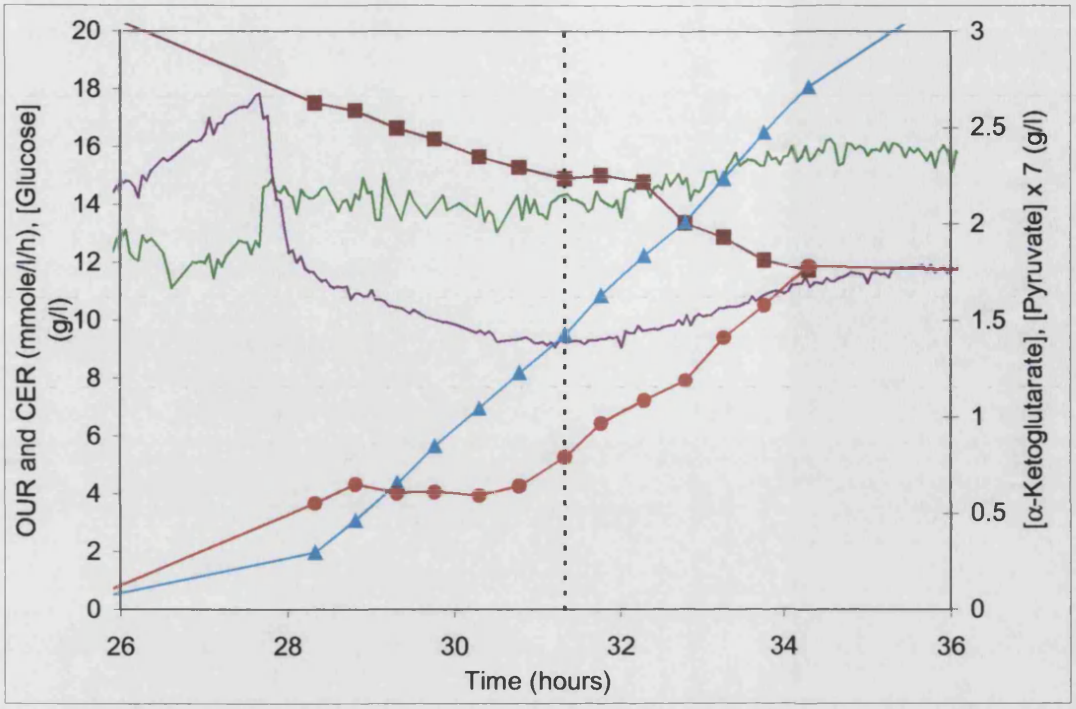


Figure 81. Uncoupler experiment control fermentation 3. This graph shows the profiles of interest around the addition time for the DMSO. Glucose ■, α-ketoglutarate ▲, pyruvate ●, OUR —, CER —.. the dashed line marks the addition of the DMSO.

This repeated unexpected behaviour in the pyruvate profile warranted further investigation. The production rate in C-mole of pyruvate and α-ketoglutarate was calculated between each sampling point in the profile. The results were plotted against each other in a bar chart (see Figure 83). The plot seems to show a dynamic response. Upon addition of the uncoupler the α-ketoglutarate secretion rate falls drastically. The pyruvate secretion however increases slightly. When α-ketoglutarate production starts again the production rate is temporarily higher than when the uncoupler was added. The pyruvate production rate on the other hand falls to almost zero.

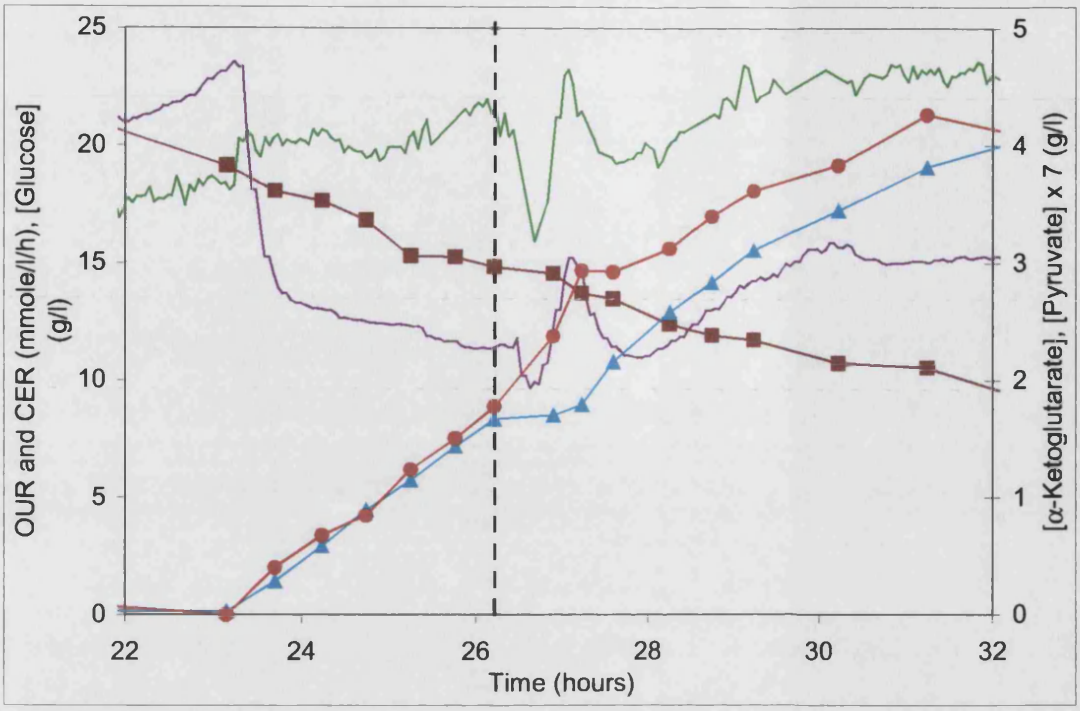


Figure 82. Uncoupler experiment, experimental 3. This graph shows the profiles of interest around the addition time for the uncoupler. Glucose ■, α-ketoglutarate ▲, pyruvate ●, OUR —, CER —. the dashed line marks the addition of the uncoupler.

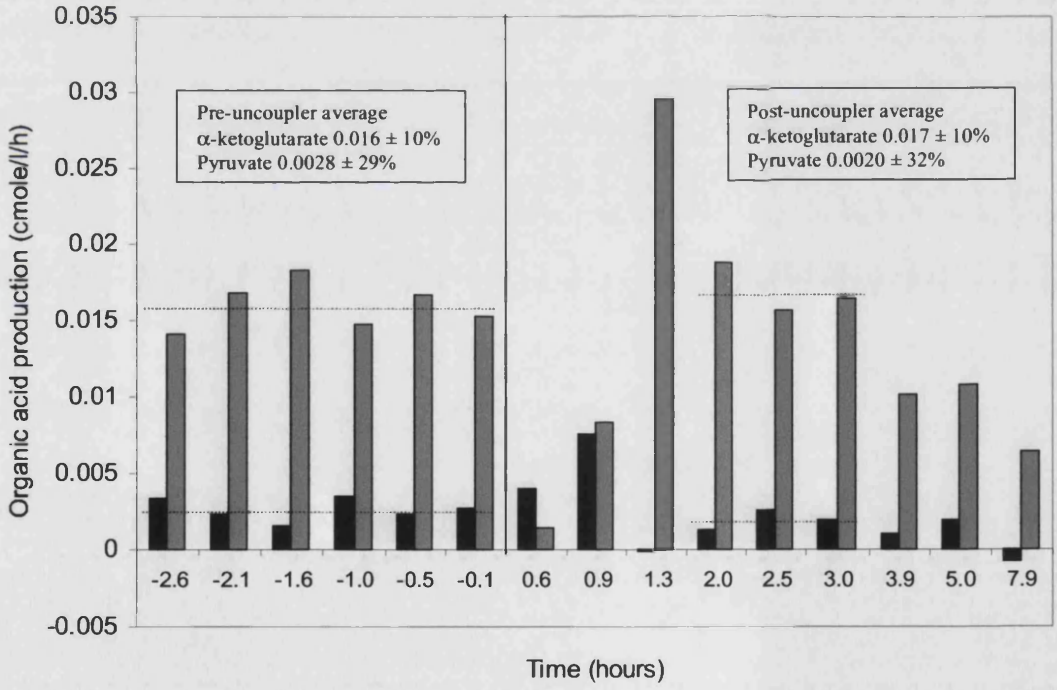


Figure 83. Uncoupler fermentation organic acid production. This graph shows the production rate of α-ketoglutarate and pyruvate at consecutive time points in the stationary phase. α-ketoglutarate ■, pyruvate ■. Times are recorded relative to the addition of the uncoupler which has been set to zero hours. The addition of the uncoupler is shown by a solid vertical line. Average values before uncoupler addition and for the first three readings after stabilisation for both pyruvate and α-ketoglutarate are marked with horizontal dashed lines.

This kind of dynamic response suggests that the uncoupler may be preventing α-ketoglutarate secretion. This might then lead to extra pyruvate secretion caused by the α-ketoglutarate build up as the cell strives to maintain a steady state. This might occur if α-ketoglutarate secretion is dependent on the membrane potential of the cell which is dissipated by the uncoupler. If this is the case the cell does not succeed in maintaining the steady state. The glucose uptake rate is unchanged and the fall in secreted carbon due to the drop in α-ketoglutarate and CO_2 production is greater than the rise in secreted carbon due to the increase in pyruvate production. This implies that either there is a new undetected secretion product or carbon is building up intracellularly.

Accumulation of carbon could be either as macromolecular storage compounds or as an increase in the metabolite pool size of the cell. The two most obvious metabolites which could build up in the cell are glucose and α-ketoglutarate. Glucose is a good candidate because it is entering the cell and if it is not utilised it will accumulate. Moreover glucose is not transportedby the PTS system in *S. erythraea* therefore phosphorylation must occur after transport and this may be inhibited by low ATP levels. If the synthesis of α-ketoglutarate is unaffected by the uncoupler but the secretion is dependent on membrane potential then α-ketoglutarate might be expected to accumulate in the cell upon addition of the uncoupler. If an intracellular storage compound were to be formed the most obvious candidate would be polysaccharides storing the excess glucose taken into the cell. Another possibility for the accumulation of carbon in the cell in response to a toxic challenge is the formation of a protective layer around the cell, a capsule. These are known from related organisms such as *M. tuberculosis*. Bacterial capsules can be made of polysaccharide (from glucose) or polyglutamate (from α-ketoglutarate). In chapter 3.3.4 *S. erythraea* was found form a capsule in the stationary phase. This was thought to be composed of polyglutamate. One possible explanation of the missing α-ketoglutarate is channelling of α-ketoglutarate to the accelerated formation of polyglutamate in order to protect the cell from the uncoupler.

6.3.3.2 Analysis of Potential Destinations for Excess Glucose During the Experiment

The options listed here for destinations for the missing carbon all come from just two areas of metabolism, glucose derived and α-ketoglutarate derived. The gas profiles can distinguish between these two options.

The CER/OUR ratio is known as the respiratory quotient (RQ). This value combined with the absolute levels of the OUR and CER gives an indication of the redox processes occurring in the cell.

If α-ketoglutarate or glutamate were the final sink for the missing carbon the reaction would be as follows. (The transamination taking α-ketoglutarate to glutamate is neutral as far as OUR and CER are concerned.)

$$C_6H_{12}O_6 + 3.5\ O_2 \rightarrow C_5O_5H_4 + CO_2 + 4H_2O$$

In this scenario the RQ is 0.29. Any extra channelling of glucose to α-ketoglutarate or glutamate would lead to a reduced RQ. Conversely any channelling of glucose away from α-ketoglutarate production to either respiration or accumulation of carbohydrate would lead to an increased RQ.

Complete combustion of glucose to CO_2 and water follows the following equation (this assumes that any NADPH produced by isocitrate dehydrogenase is subsequently converted to NADH by transhydrogenase activity as seems to be the case in the stationary phase, see section 4.3.4.1.2).

$$C_6H_{12}O_6 + 6O_2 \rightarrow 6CO_2 + 6H_2O$$

This scenario gives an RQ value of 1. Addition of uncoupler is expected to lead to more carbon being channelled to respiration to provide more energy for the cell. In this case the RQ can be expected to move towards 1. The absolute levels of the OUR and CER will both rise as more glucose is used in this way producing more CO_2 and consuming more O_2.

If glucose were being redirected to polysaccharide formation (neglecting branching and activation by nucleoside triphosphates) the reaction would probably be of the form

$$nC_6H_{12}O_6P \rightarrow (C_6H_{10}O_5)n + n\text{-}1\ H_2O + nPi$$

This scenario has no direct requirement for O_2 or production of CO_2 and so would not alter the RQ value. The channelling of carbon away from respiration would however reduce the total amount of CO_2 produced and O_2 required. If the carbon channelled to polysaccharide formation meant that less was going to α-ketoglutarate production both the CER and the OUR would be expected to fall. The OUR would be expected to fall more than the CER because 3.5 times more O_2 is required for the production of α-ketoglutarate than CO_2 is produced. The same arguments would be true if glucose were accumulating in the cell.

In all experiments the OUR, CER and α-ketoglutarate secretion fell on addition of the uncoupler. This points to carbon being accumulated within the cell. The OUR fell considerably more than the CER. This increase in the RQ must be the result of cessation of the biosynthesis of α-ketoglutarate. This implies that the lost carbon is not being accumulated as α-ketoglutarate or glutamate. Furthermore it implies that the excess glucose is not being catabolised through either glycolysis or the pentose phosphate pathway because this would effect the OUR and CER readings.

Twenty to thirty minutes after the fall in the OUR and CER profiles the trend changes and the values exceed those seen before the addition of the uncoupler (see Figure 82). The size of the reduction was bigger for the OUR than the CER, but the sizes of the increases seem to be roughly equal. This implies that there is extra respiratory activity for a short period before activity returns to pre-uncoupled levels. This increase reflects the fact that the cell is adapting to the uncoupler and readjusts its metabolism. It may be that the overshoot seen in the OUR and CER readings is due to the persistence of uncoupler activity. Alternatively the cell may be recharging the adenylate pool and the reduced cofactor pool. It has been noted in chapter 4 that *S. erythraea* wastes large amounts of energy in the stationary phase. If the cell were to reduce this wastage it

might be able to accommodate the wastage caused by the uncoupler. This may explain how the cell recovers from addition of the uncoupler so quickly.

6.3.3.3 Refinement of the Hypothesis

It is possible from this analysis to form a new hypothesis as to what might be happening in the cell. Before addition of the uncoupler glucose is taken into the cell and passes through glycolysis with some flux splitting off through phosphoenolpyruvate carboxylase in order to be combined with acetyl-CoA to form citrate and eventually α-ketoglutarate for secretion. The rest of the acetyl-CoA is used to drive respiration. Upon addition of the uncoupler the production of ATP is slowed and the adenylate energy charge of the cell drops. In *S. erythraea* glucose is not thought to be imported by the PTS system which is driven by energy from hydrolysis of the phosphate bond in the phosphorylated glycolysis intermediate phosphoenolpyruvate. Rather glucose is first taken into the cell and subsequently phosphorylated by ATP. As glycolysis proceeds further phosphorylation occurs requiring more ATP (see Figure 84). The drop in the adenylate energy charge in the cell upon addition of the uncoupler may restrict the flux through glycolysis leading to a build up of glucose and possibly other sugar compounds. In this respect Hockenhull *et al.* (1954) found that inhibition of glycolysis using sodium fluoride in *Streptomyces griseus* led to the accumulation of glucose-1-phosphate, glucose-6-phosphate and a hexose diphosphate, possibly fructose-1-6-bisphosphate. The reduced flux through the early steps of glycolysis would prevent overflow metabolism stopping the secretion of α-ketoglutarate and reducing the activity of the TCA cycle. This would further restrict availability of ATP for the early steps of glycolysis. The OUR and the CER would both fall but the OUR would fall the most due to the ratios of O_2 required and CO_2 produced by α-ketoglutarate synthesis.

The cell recovers from its initial shock on addition of the uncoupler. As the cell recovers it contains excess sugars and has a low adenylate energy charge and a high ratio of NAD^+ to NADH. Low levels of the uncoupler are still present however leading to a lower efficiency of oxidative phosphorylation. This leads to higher TCA cycle activity and higher OUR and CER values. The extra flux through the TCA cycle leaves no carbon for α-ketoglutarate secretion. Once the adenylate energy charge and the

NAD^+ to NADH ratio return to normal levels the remaining excess sugars are processed and secreted as α-ketoglutarate leading to a temporary higher flux of α-ketoglutarate. Eventually the cells return to the same metabolic pattern displayed before the addition of the uncoupler.

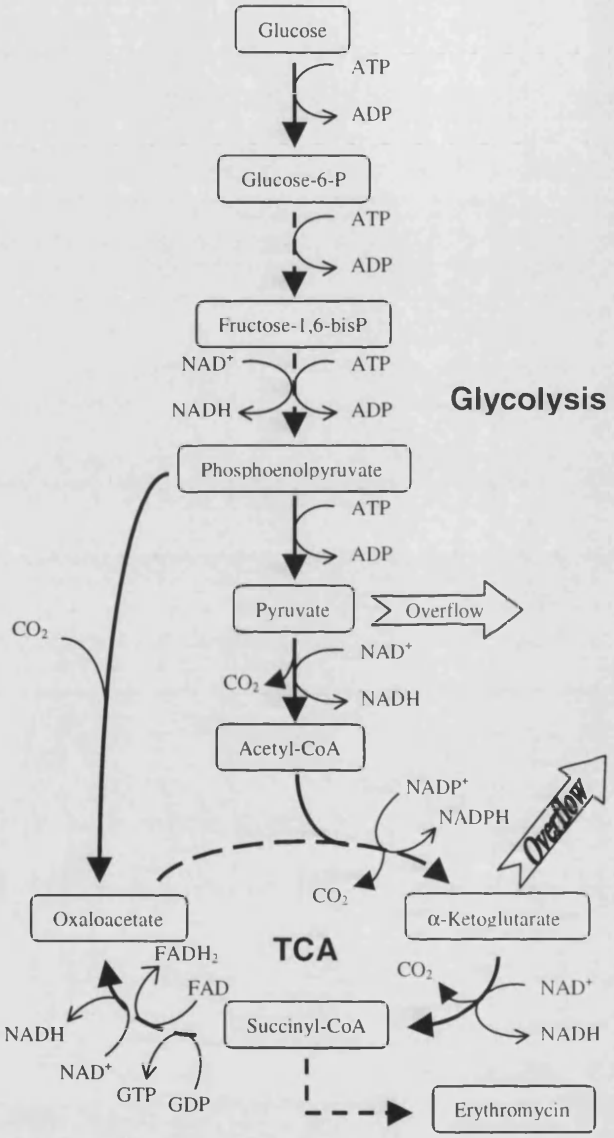


Figure 84. Central metabolism in *S. erythraea*

In this interpretation of the results the initial hypothesis is maintained. The predicted increase in OUR and CER on addition of the uncoupler is not seen immediately because the cells undergo a large metabolic shock and they are unable to process all the glucose taken up through glycolysis. This masks the effect on the OUR and CER until the cells recover from this shock whereupon the effect is seen.

Although this hypothesis explains some of the data it does not explain the pyruvate profile. It is not particularly clear why pyruvate should continue to be secreted whilst α-ketoglutarate secretion is halted. An argument could be made that cessation of the α-ketoglutarate secretion primarily reduces the anaplerotic flux from phosphoenolpyruvate to oxaloacetate leaving the flux from phosphoenolpyruvate to acetyl-CoA less changed thus preventing changes to the pyruvate profile. This could also explain the two different pyruvate profiles observed in the first fermentation. If the flux through glycolysis falls below a certain threshold reduced pyruvate secretion would be observed. Above this threshold there would be no observable change in pyruvate secretion because of buffering by loss of α-ketoglutarate production. This argument seems weak for three reasons. Firstly it predicts that at higher uncoupler concentrations the pyruvate profile will follow the same profile as the α-ketoglutarate profile, however the opposite is seen. In the shake flask experiments at lower concentrations the profiles of α-ketoglutarate and pyruvate seem more similar than at high concentrations. Secondly, in all the reports of organic acid production in the *streptomyces* pyruvate is the main secreted acid and α-ketoglutarate only seems to be secreted when high levels of pyruvate secretion are already occurring (see section 1.5.3.2.2). Thirdly this argument cannot explain the delayed pause in pyruvate secretion.

The hypothesis formulated here predicts that there will be significant accumulation of carbon in the cell as carbohydrate, probably glucose and that accumulation of α-ketoglutarate will not be seen. To test this hypothesis the experiment was repeated and samples taken for analysis of intracellular α-ketoglutarate, glucose and pyruvate.

6.3.4 Repeat of Uncoupler experiment with Analysis of Intracellular Metabolites

In this experiment duplicate control and experimental fermentations were performed, for brevity only results from control fermentation 4 and uncoupler fermentation 4 are presented. Results from the other two fermenters were in agreement with these two. The fermentation profiles are presented as Figure 85 and Figure 86.

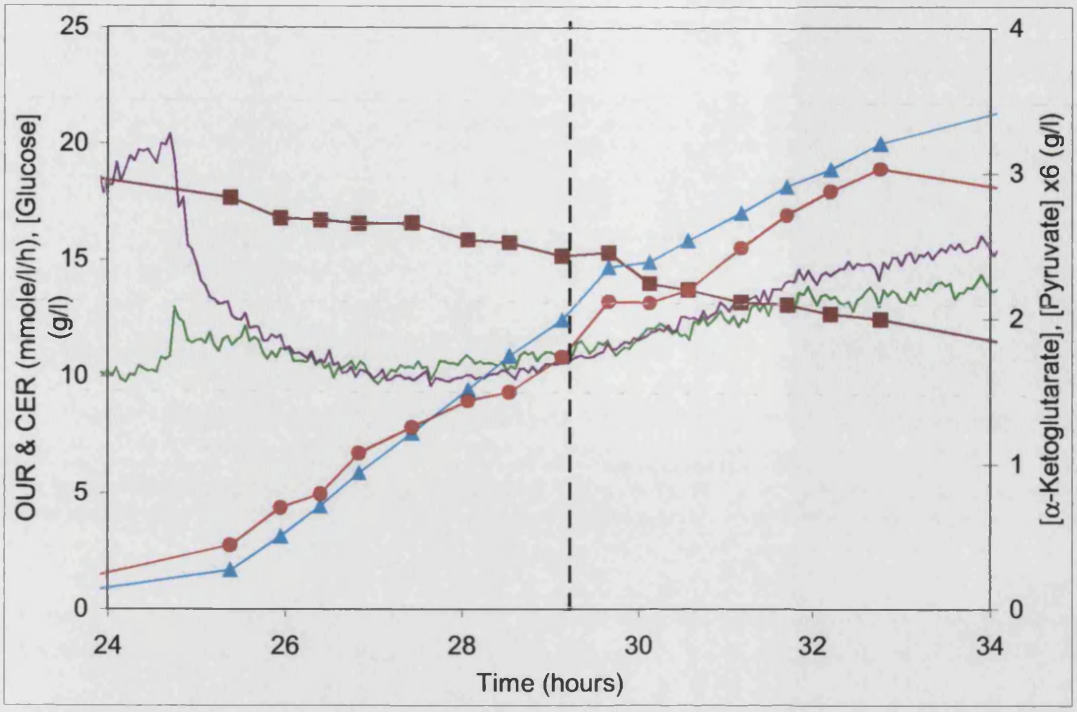


Figure 85. Uncoupler experiment control fermentation 4. This graph shows the profiles of interest around the addition time for the DMSO. Glucose ■, α-ketoglutarate ▲, pyruvate ●, OUR —, CER —. The dashed line marks the addition of DMSO. Pyruvate concentrations have been multiplied by 6 in order to facilitate comparison with the α-ketoglutarate concentrations.

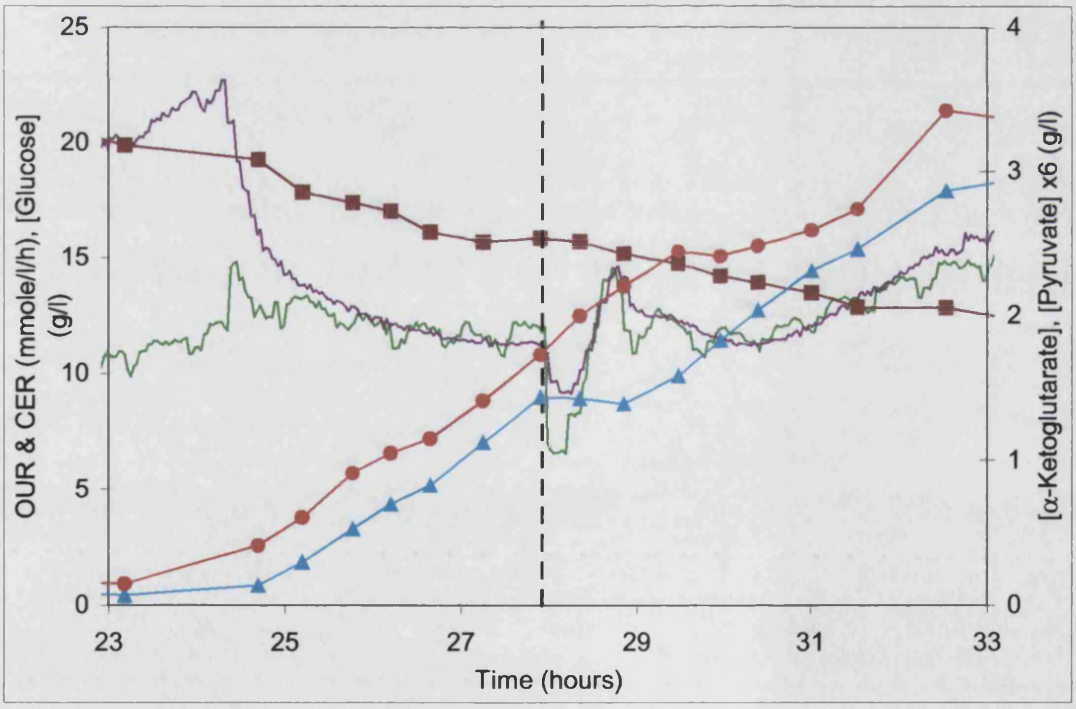


Figure 86. Uncoupler experiment, experimental fermentation 4. This graph shows the profiles of interest around the addition time for the uncoupler. Glucose ■, α-ketoglutarate ▲, pyruvate ●, OUR —, CER —. The dashed line marks the addition of DMSO. Pyruvate concentrations have been multiplied by 6 in order to facilitate comparison with the α-ketoglutarate concentrations.

The fermentation profiles show good similarity with previous experiments. The α-ketoglutarate, pyruvate, glucose and off gas profiles support the previous finding that α-ketoglutarate and pyruvate production pauses are separated. Once again however the carbon balance does not close following the addition of the uncoupler indicating an unmeasured carbon sink.

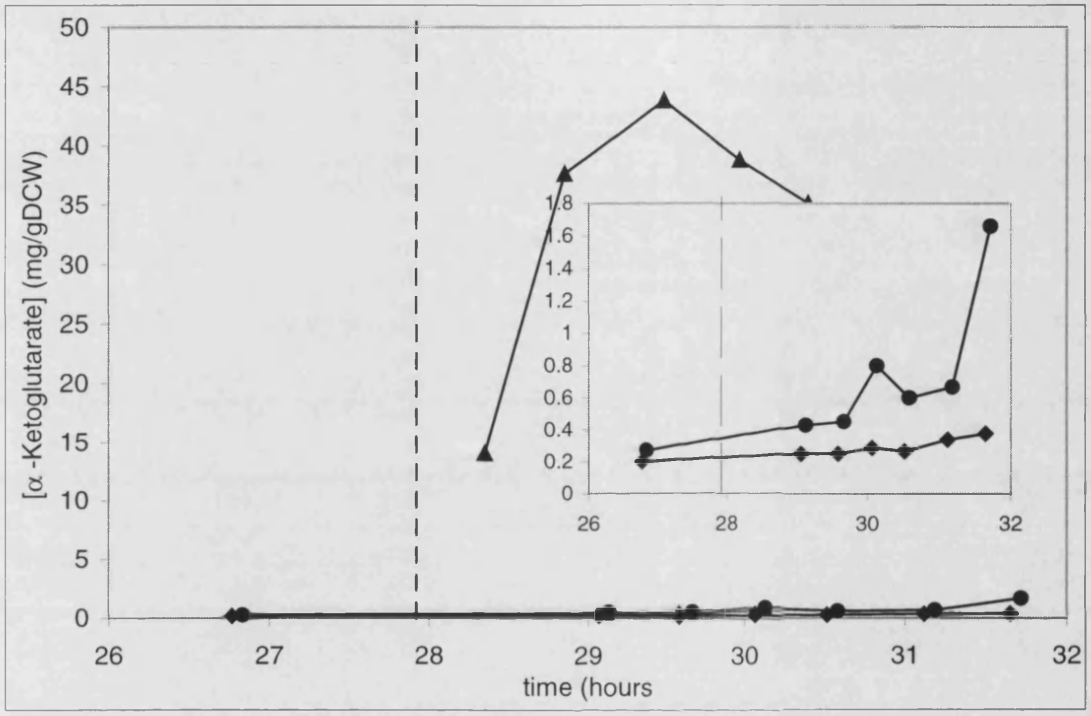


Figure 87. Accumulation of intracellular α-ketoglutarate. This graph shows the actual and predicted intracellular accumulation of α-ketoglutarate on addition of uncoupler. The accumulation in the control is also shown. The inset graph is a magnification of the uncoupler and the control. Uncoupler ◆, control ●, predicted ▲. Addition of uncoupler is marked by a dashed line.

In order to investigate the possibility of the accumulation of α-ketoglutarate, glucose and pyruvate intracellular levels of these compounds were measured. The pyruvate levels were below the limit of detection. Results for the α-ketoglutarate and glucose levels can be seen in Figure 87 and Figure 88. By calculating the rate of production of α-ketoglutarate before and after addition of the uncoupler it was possible to estimate how much α-ketoglutarate would have accumulated in the cell if it did not overflow elsewhere. This predicted profile has also been included on the graph. The levels of α-ketoglutarate measured in the cells did not differ significantly between the control and the uncoupler experiments. The levels found were approximately 1% of those predicted. From this it was concluded that, in agreement with the hypothesis, α-ketoglutarate does not accumulate in the cell upon addition of the uncoupler. Likewise

by performing a carbon balance across the reactor the amount of glucose that would have accumulated was calculated. This predicted profile has also been included. The levels of glucose measured did not differ significantly between the control and the uncoupler experiments and are much lower than the predicted levels. From this it was concluded that, contrary to the hypothesis, glucose does not accumulate in the cell upon addition of the uncoupler.

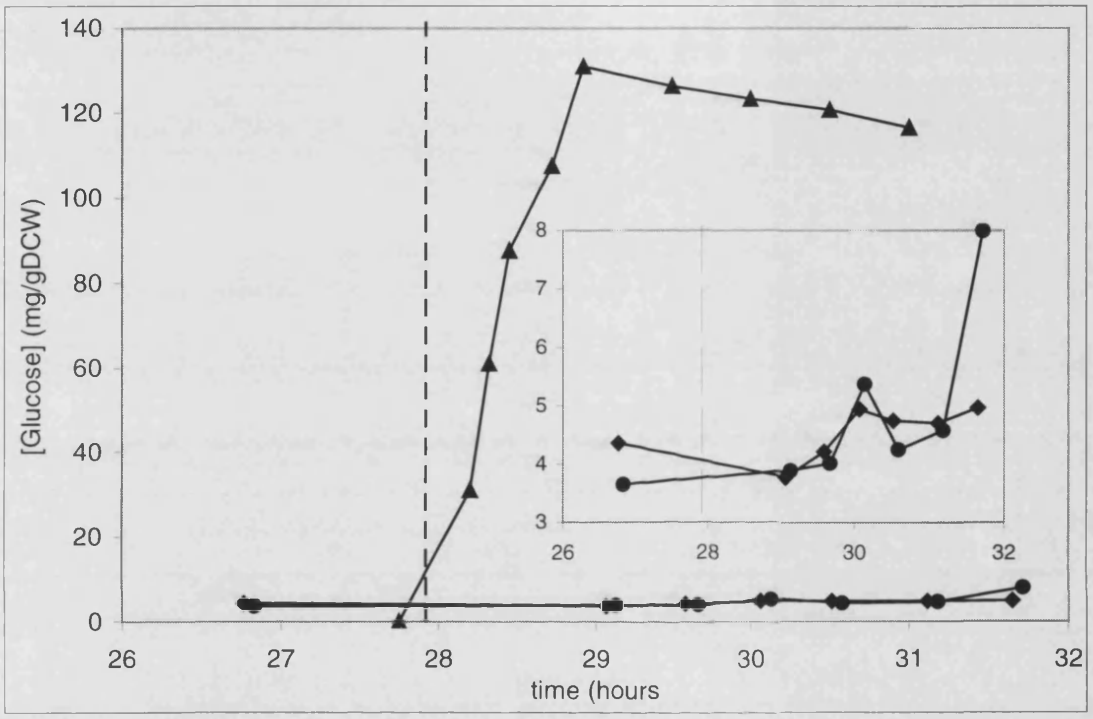


Figure 88. Accumulation of intracellular glucose. This graph shows the actual and predicted intracellular accumulation of glucose on addition of uncoupler. The accumulation in the control is also shown. The inset graph is a magnification of the uncoupler and the control. Uncoupler ◆, control ●, predicted ▲. Addition of uncoupler is marked by a dashed line.

This raises the question as to where the carbon entering the cell is going. It has not been found as an extra cellular product and it does not seem to be as any of the most obvious intracellular metabolites. It is possible that the glucose is being converted into a sugar storage compound. Trehalose is accumulating in the cell during this stage of the stationary phase (see section 3.3.3.2). It could be that the excess glucose taken up is converted to trehalose however this would require activation of the glucose by nucleoside triphosphates. These are likely to be in short supply in a cell with low adenylate energy charge. Alternatively the extra carbon could be distributed throughout the whole metabolite pool of the cell. This seems unlikely though as it implies some form of balanced metabolism rather than the drastic dynamic response suggested by the off gas profiles. The next step in this work should probably be to test intracellular

trehalose, or possibly more generally intracellular carbohydrate during uncoupler addition. Inspection of the carbon balance revealed that the excess carbon taken up by the cell on addition of the uncoupler was not quickly utilised when the cell recovered. This suggestes that it is sequestered in some form of storage compound rather than in metabolic intermediates.

This work seems to show that the uncoupler interferes with glycolysis at the 10 µmole/l used in this experiment. The effect on organic acid production is however little changed from the first shake flask experiment where only a 2.5 µmole/l was used. It could be that the high concentration of uncoupler is responsible for the effect on glycolysis. It might be instructive to repeat the experiment with lower concentrations to see if the same effects are seen.

The failure to close the carbon balance in these experiments led to a renewed effort to identify extracellular products. During the course of these experiments one such product was identified (see section 4.3.3). Fumaric acid, another organic acid from central metabolism, was found to accumulate in the media during the stationary phase in a manner similar to pyruvate and α-ketoglutarate. The fumarate levels during this last set of fermentations were recorded and the profiles are presented in Figure 89. The profiles seem to show another slightly different trend upon addition of uncoupler. There is an immediate very short pause in production followed by an acceleration in fumarate production followed by another pause before production goes back to the values seen before addition. The levels of fumarate produced were significantly lower than the levels of pyruvate hence it did not account for all the missing carbon from the carbon balance. These results like the pyruvate results are not easily reconciled with the original hypothesis formed as to organic acid secretion.

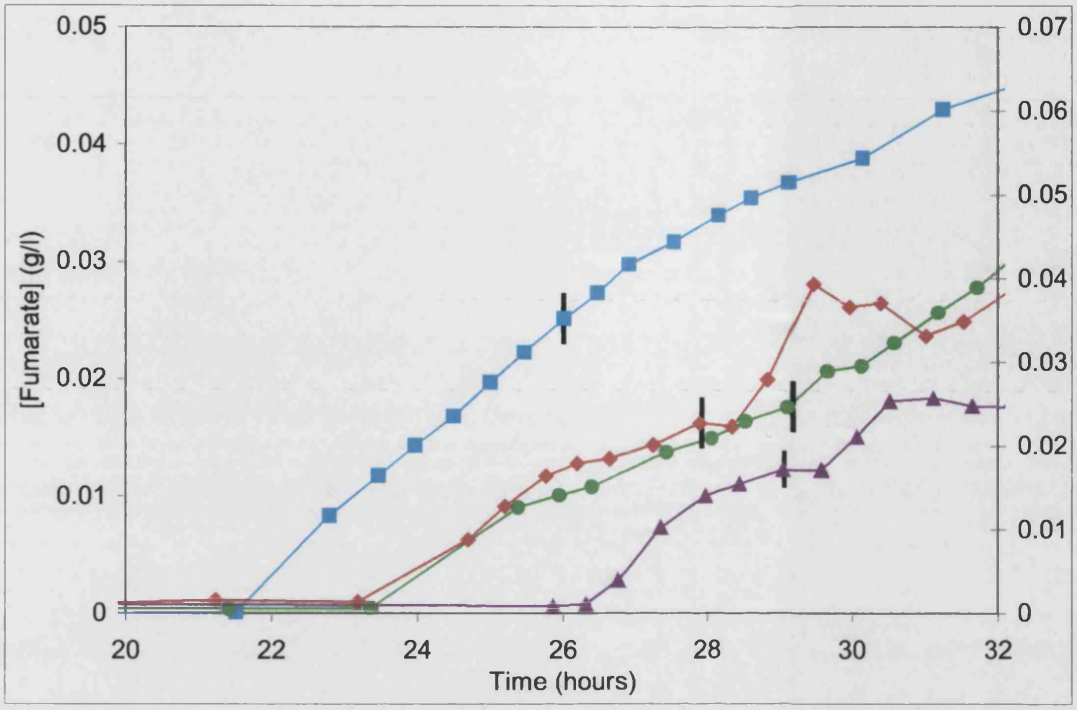


Figure 89. Fumarate profiles from control and experimental fermentations 4 and 5. This graph shows the fumarate secretion profiles for the four bioreactors used in the third uncoupler fermentation experiment. Control fermentation 5 ■, uncoupler fermentation 5 ▲, control fermentation 4 ●, uncoupler fermentation 4 ◆. Addition of uncoupler or DMSO is marked on each profile by a vertical bar.

6.3.5 Implications for Network Flexibility

The findings in this chapter also have implications for the flexibility of metabolism. Absolute values of fluxes have not been calculated by FBA due to inability to close the carbon balance in these experiments. However a number of inferences regarding fluxes in the pathways of central metabolism have been made from the data. The results seem to demonstrate that flux round the TCA cycle is rigidly tied to the flux through oxidative phosphorylation. The implications of this for erythromycin synthesis are discussed in the conclusions section. The work also however sheds light on two other key areas of metabolism, the branch points at glucose-6-phosphate and phosphoenolpyruvate which were discussed in section 4.3.4.5.

The addition of uncoupler has an immediate impact on α-ketoglutarate secretion. α-Ketoglutarate secretion is supplied by anaplerotic flux through phosphoenolpyruvate carboxylase. Addition of the uncoupler halts the flux through phosphoenolpyruvate carboxylase whilst the flux through pyruvate kinase continues only slightly reduced.

This shows that the flux split at phosphoenolpyruvate is not rigidly determined, the node is flexible, this agrees well with the findings in section 4.3.4.5.2.

If the hypothesis that glucose accumulating after the addition of the uncoupler is converted into trehalose is correct then the glucose-6-phosphate branch point must also be flexible. The organism is able to re-channel flux from glycolysis to the synthesis of glucose-1-phosphate on addition of the uncoupler. If this is so it agrees with the findings in section 4.3.4.5.3.

The uncoupler experiment confirms that the flexibility seen in the metabolic network of *S. erythraea* is not a property of changing isozyme usage during the course of the fermentation. Rather it is derived from the properties of the individual enzymes. The flexibility of key branch points during the stationary phase has important implications for metabolic engineering of the organism. Flexible branch points mean that the network should be amenable to supporting changes in the flux distribution introduced by metabolic engineering.

The flexibility of the metabolic network may have implications for the organic acids secreted by *S. erythraea*. One of the differences found in organic acid secretion between *S. erythraea* and the *streptomyces* species reported in the literature is that the main organic acid secreted by *S. erythraea* is α-ketoglutarate whereas in *streptomyces* species it seems to be pyruvate. In section 4.3.4.1.2 it was noted that the TCA cycle activity found in the stationary phase would produce more NADPH than the cell had capacity to reoxidise. One of the proposed mechanisms by which the cell might alleviate this was the use of a pathway found to operate in *E. coli* by Fischer *et al.* (2003). This pathway avoids isocitrate dehydrogenase by utilising isocitrate lyase bypassing α-ketoglutarate and succinyl-CoA. If this pathway does operate in *S. erythraea* during the stationary phase some regeneration of NADPH will still be required. The secretion of α-ketoglutarate may be a result of the cells need to reduce $NADP^+$ through low levels of isocitrate dehydrogenase activity. This would produce α-ketoglutarate. If the full TCA cycle is not operating then there may be no way to further oxidise this α-ketoglutarate leading to its secretion. This might explain why α-

ketoglutarate is secreted in larger amounts than pyruvate in stationary phase *S. erythraea.*

6.4 Conclusions

The hypothesis that organic acid secretion is due to reduced levels of NAD^+ in the cell during the stationary phase developed in chapter 4 is given a cautious approval by these results. It seems likely that the initial pause in α-ketoglutarate production seen on addition of the uncoupler is caused by effects earlier in metabolism rather than the portion of metabolism being tested. This masks the effect that is being looked for. However as the cell adapts to the situation the predicted effect is seen as carbon is channelled away from α-ketoglutarate production towards respiration. This indicates that TCA cycle flux is limited primarily by the demands of respiration. The off gas profiles associated with the early response to uncoupler addition led to a new hypothesis. This hypothesis was that the requirement for ATP in the early stages of glycolysis, at a time when the uncoupler had reduced the levels of ATP, meant that carbon was being stored in the cell as carbohydrate. Tests confirmed that the carbon accumulating was not accumulating as α-ketoglutarate or pyruvate but also suggested that it was not accumulating as glucose. Other potential carbohydrate candidates include glycogen and trehalose. Trehalose is favoured as it is already accumulating substantially in the cell during the stationary phase. The levels of accumulation suggested here are about 12% of the DCW. This is similar to the levels seen through normal accumulation. This however would be on top of normal accumulation and occurs much more quickly.

Further work possible in this area would includes repetition of the fermentations with lower levels of uncoupler to see if the masking effect suggested above can be removed giving a clearer indication of the link between demand of reduced cofactors for respiration and flux through the TCA cycle. The levels of trehalose could also be measured during a fermentation to investigate further the fate of the unaccounted carbon.

One of the unexpected findings of this work was that the cells recover very quickly from the addition of the uncoupler. An hypothesis was proposed that the cell was able

to recover by increasing the efficiency of its use of energy so as to negate the loss of energy due to the addition of the uncoupler. This was considered possible because of the large apparent wastage of energy seen in the stationary phase (see section 4.3.4.3). The implications of this work for the overproduction of erythromycin or other methylmalonyl-CoA derived compounds are complicated. It shows that flux round the TCA cycle is firmly controlled by respiratory requirements. However it is of no practical use to increase flux by increasing respiratory requirements as done here. This merely leads to more carbon being converted to CO_2 and less carbon leaving the TCA cycle to form products, waste or desired. Rather two aspects of metabolism will have to be addressed to overcome these limitations. There will have to be a greater demand, drawing TCA cycle intermediates out of the cycle at the required point. Secondly this demand will have to be supplied without altering the balance of reduced cofactor production in the cell because the TCA cycle is strongly regulated to maintain the cofactor balances. If such changes are made the cell may act to restore these balances within the framework imposed by such changes. Alternatively it could act to restore these balances in a way that negates the changes introduced. The fixed flux through the TCA cycle in the stationary phase may work to the metabolic engineer's advantage. It guarantees that at high glucose uptake rates there will be significant overflow from the TCA cycle which is what is needed for erythromycin overproduction.

Chapter 4 showed that maximum yields of erythromycin will be achieved by operating with a split TCA cycle. This can only be done by decoupling flux through oxidative phosphorylation from flux through the TCA cycle. The NADH requirements of oxidative phosphorylation will have to be supplied from other sources such as the pentose phosphate pathway. In this case the limit imposed on TCA cycle flux by the NAD^+ to NADH ratio and the adenylate charge might work in the favour of the metabolic engineer by shutting down the TCA cycle at α-ketoglutarate allowing it to operate in a split fashion.

This chapter has demonstrated the potential of uncouplers of oxidative phosphorylation for use in investigating metabolic processes relevant to industrial fermentations. However it has also shown that the effects these compounds have may be complex and

difficult to interpret. Alongside uncouplers other compounds might be useful in these studies such as inhibitors of the electron transport chain or of other areas of metabolism.

7 GENERAL DISCUSSION AND CONCLUSIONS

The aim of this project was to apply flux balance analysis to the secondary metabolism of *S. erythraea*. In order to accomplish this the peculiar physiology of actinomycete organisms had to be taken into account. Once established this method was to be used to investigate the secondary metabolism of *S. erythraea*. In this section the findings will be summarised in a number of key areas. These will be: the impact of biomass composition on modelling, findings concerning the rigidity of the metabolic network, the wastage of carbon and energy by *S. erythraea* and the overproduction of erythromycin. Some possible future directions for the work will then be discussed.

7.1 Biomass composition

During the course of this project a number of new findings regarding the biomass composition of *S. erythraea* have been made. The existence of a capsule produced in the stationary phase has been established (see section 3.3.4). It is thought likely that this is composed of polyglutamate as seen in the closely related *Mycobacteria* (Harth *et al.* 2000). Trehalose has been reported in *S. erythraea* for the first time and is found to be synthesised both in the growth phase and the stationary phase (see section 2.6.6). When glucose is used as carbon source trehalose accumulates in the stationary phase under nitrogen limitation to around 15% of the biomass, it is then reconsumed when glucose becomes limiting. The contribution of arabinogalactan to the biomass has been determined for the first time, it forms a significant proportion of the biomass, around 15% throughout batch culture. The monomeric composition of the macromolecules of *S. erythraea* have been estimated for the first time (see section 3.3.2). This however is strongly dependent on information from other species and is likely to become obsolete soon with the forthcoming publication of the *S. erythraea* genome (McKay *et al.* 2001).

In chapter 3 the biomass composition of *S. erythraea* was shown to be dynamic, changing substantially during the course of the fermentation like in the closely related *streptomyces* (see King and Büdenbender, 1997 and Olukoshi and Packter, 1994). The

composition is reasonably constant during the growth phase, however as the fermentation enters the stationary phase the composition alters significantly. This is particularly noticeable in two groups of biomass components. Nitrogen containing components show large changes when the nitrogen source becomes depleted. Protein and RNA levels fall from the onset of stationary phase, presumably linked to lower requirements for their activity once rapid growth has ceased (Shahab *et al.* 1996). As the protein and RNA decrease a new component, the capsule, which is thought to contain nitrogen is synthesised. A second group of biomass components also shows significant dynamic behaviour. Under conditions of nitrogen limitation carbon storage compounds accumulate in the biomass. This is seen in the accumulation of both lipid and trehalose in the stationary phase, the trehalose is also reused once the glucose runs out.

The dynamic nature of the biomass composition of *S. erythraea* raised questions about the validity of some of the assumptions normally made when performing flux balance analysis on microorganisms. A fixed biomass composition is usually assumed and this does not take into account dynamic patterns in the biomass composition (Stephanopoulos *et al.* 1998). Flux balance analysis was performed for batch fermentations of *S. erythraea* at a variety of time points throughout the fermentation. Analysis was performed twice, once assuming fixed biomass composition and once taking into account the dynamic nature of the biomass composition. During the exponential phase when biomass composition is constant there was little difference between the methods. In the stationary phase however there were significant differences. Before the glucose became depleted these differences were generally only large outside the core pathways of glycolysis, TCA cycle, respiration and the pentose phosphate pathway. After depletion of glucose large discrepancies between the flux distributions calculated by the two methods were found in all areas of metabolism. The differences found were considered to be significant enough to warrant measurement of biomass composition when performing flux balance analysis of *S. erythraea* especially if peripheral pathways such as pathways of secondary metabolism are of interest (see section 3.3.5).

7.2 Network Rigidity

Changes in the flux distribution in *S. erythraea* during the course of the fermentation were studied in chapter 4. There are very large changes to the distribution of flux as the fermentation moves from growth phase to stationary phase. These changes are seen both in the central pathways and in the peripheral pathways. Further radical changes in the flux distributions are seen when the glucose becomes depleted. This suggested a great flexibility in the metabolic network of *S. erythraea* enabling it to radically alter its metabolism between phases. A closer inspection revealed that the key branch points in central metabolism displayed a very high degree of flexibility over the course of the fermentation. This suggests that metabolic engineering of the central metabolism of *S. erythraea* should not be hampered by tight regulation of these branch points. However it is also possible that some of this apparent flexibility is due to expression of different isozymes of key enzymes at different stages of the fermentation. The flexibility of the metabolism of other actinomycetes has not been reported in this way before. However van Gulik *et al.* (2000) have performed this kind of analysis on *Penicillium chrysogenum* and found that the principle nodes in this organism were flexible.

The work performed in chapter 6 with the uncoupler supports the proposed flexibility of the metabolic network. One of the key nodes in the network is glucose-6-phosphate. Addition of an uncoupler to stationary phase cells greatly reduced the flow of carbon through glycolysis but did not slow glucose uptake. This suggests that glucose uptake in *S. erythraea* is not dependent on the membrane potential of the cell unlike glucose uptake in *Streptomyces coelicolor* (Bertram *et al.* 2004). Glucose did not accumulate in the biomass leading to the suggestion that the glucose was stored as trehalose. If this is the case it demonstrates that the glucose-6-phosphate node is truly flexible during the stationary phase allowing redistribution of the flux from the pathways proceeding from it instantaneously. Moreover the flux split ratio at the phosphoenolpyruvate branch point also changed upon addition of the uncoupler. The secretion of α-ketoglutarate from the cell requires a high anaplerotic flux in order to maintain levels of TCA cycle intermediates. Results in chapter 6 indicated that this flux stopped upon addition of the uncoupler whilst the flux through pyruvate to citrate although slightly reduced did not stop. This finding suggests that the flux split ratio at this branch point is not rigidly

controlled either because the flux through one branch can change independently of the flux through the other branch.

Although the main branch points of metabolism were found to be flexible there are some areas of metabolism in *S. erythraea* where rigid constraints do occur. One such area is the TCA cycle. Experiments performed in chapter 6 with an uncoupler suggest that the flux through the TCA cycle is closely related to the requirement of NADH for oxidative phosphorylation. This means that there is little leeway for increasing flux through the TCA cycle. However increasing flux around the TCA cycle is of little commercial interest as this is merely increasing production of CO_2. The real interest is in increasing flux across the TCA cycle, providing more flux out of the TCA cycle via its intermediates and this requires increased anaplerotic flux. In the wildtype strain flux round the TCA cycle is adequate to supply considerably more flux across the TCA cycle than found here. Given this capacity the rigidity of this aspect of the metabolic network is probably not important even when considering substantial increases in flux from the TCA cycle to erythromycin biosynthesis. If maximum theoretical yields of erythromycin are required however the TCA cycle will have to operate in a split mode with parts running in the reverse of the usual direction (see section 4.3.6). This will only be able to occur if the cells NADH requirements for oxidative phosphorylation are met from elsewhere within the metabolic network.

The balancing of NADPH production and consumption is another rigid constraint that would be expected to influence the flux distribution. During the stationary phase the demand for NADPH is low. The TCA cycle of *S. erythraea* produces both NADH and NADPH. The NADH requirements of the cell are such that TCA cycle flux would produce more NADPH than the cell requires. This would lead to unbalanced and therefore unsustainable metabolism. The actual metabolic network of *S. erythraea* must have the in built flexibility to avoid this. The model does not reflect this. Two potential mechanisms are proposed for this flexibility: a pathway recently demonstrated in *E. coli* by Fischer and Sauer (2003), which bypasses isocitrate dehydrogenase, and a futile cycle utilising malic enzyme which acts as a transhydrogenase. Both solutions make use of some of the several anaplerotic pathways in *S. erythraea*. These can be coupled in a large number of combinations to achieve different metabolic effects. This kind of innate flexibility which is build into metabolic networks may be very important in

organisms such as *S. erythraea* which display large metabolic changes between different stages of their growth.

7.3 Wastage of Carbon and Energy

One of the aspects of *S. erythraea* metabolism which seems interesting is its use of carbon and energy, especially in the stationary phase. The wild type organism produces large quantities of organic acids during the nitrogen limited stationary phase (see chapter 4). Organic acid production has been reported in a number of *streptomyces* species especially pyruvate and α-ketoglutarate (see section 1.5.3.2.2). These organic acids were found in *S. erythraea* along with fumaric acid. Fumaric acid production has not been reported in *S. erythraea* previously and I am unaware of it being reported for any other actinomycete. Levels of secretion were lower than for α-ketoglutarate and pyruvate. All these organic acids are reused later in the fermentation, however their overproduction suggests that the metabolism of *S. erythraea* is not adjusted for experiencing high glucose under conditions where biomass synthesis has stopped.

Another aspect of this apparently wasteful metabolism is that the level of oxidative phosphorylation appears not to be directly linked to the metabolic requirements of the cell (see section 4.3.4.3). These maintenance energy requirements are expected to be reasonably constant (Pirt 1965), however in *S. erythraea* they seem to vary greatly according to the conditions experienced by the cell. Under conditions of excess glucose very high maintenance associated oxidative phosphorylation is observed, however when glucose is depleted the maintenance requirements can fall to less than 20% of their peak levels. This suggests a very inefficient use of energy during times of glucose excess. This inefficient use of energy may explain the transitory effect of uncoupler addition to stationary phase *S. erythraea*. By increasing the efficiency of its energy usage the cell may be able to make up for the inefficacy in oxidative phosphorylation cause by the uncoupler.

The wastage of carbon and energy can be linked to another observation. The uptake of glucose by *S. erythraea* under conditions of excess glucose does not seem to be tightly

linked to the demand for carbon and energy in the cell (see section 4.3.4.2.3). Wild type *S. erythraea* maintains high levels of glucose uptake in the stationary phase even though no growth is occurring. Moreover the industrial overproducing strain CA340 displays a much lower glucose uptake rate in the stationary phase with lower secretion of organic acids and lower maintenance energy utilisation. It was not clear however whether the production of organic acids and the inefficient use of carbon caused the high levels of uptake of glucose, or whether the uncontrolled uptake of glucose led to leakage of organic acids and energy by the cell. It has been possible to determine this for the stationary phase of *S. erythraea* using an uncoupler of oxidative phosphorylation. Addition of uncoupler to stationary phase cells caused a large reduction in organic acid secretion and in oxidative phosphorylation, however the uptake of glucose continued at the same rate as before the addition. Analysis of the off gas data revealed that the glucose was probably not being processed through glycolysis or the pentose phosphate pathway. It was proposed that the glucose accumulated within the cell probably as trehalose. This continued uptake of glucose in the absence of demand for glucose in the cell demonstrates that the glucose uptake is independent of the cell's demand for glucose. This strongly suggests that it is the uncontrolled uptake of glucose that drives the wasteful use of both carbon and energy in the stationary phase. This conclusion also aggress with the findings of Surowitz and Pfister (1985). They found that pyruvate secretion in *Streptomyces alboniger* was caused by an imbalance in the enzymes of glycolysis compared to those of the TCA cycle.

The lack of regulation of glucose uptake alone cannot explain the production of organic acids, this also requires limitation of the organism's ability to process the carbon taken into the cell. This limitation in the stationary phase can be attributed to the cessation of biomass production which reduces the demand for carbon and energy from central metabolism. This may explain the cause of organic acid secretion at a general level but does not give the specific details. In chapter 6 TCA cycle activity was investigated to determine whether activity is limited by the reoxidation of NADH in oxidative phosphorylation. The evidence found whilst not conclusive suggests that it may be. This limitation would take effect at pyruvate dehydrogenase and α-ketoglutarate dehydrogenase leading to secretion of pyruvate and α-ketoglutarate as observed in stationary phase *S. erythraea*. An uncoupler of oxidative phosphorylation was added to

stationary phase cells of *S. erythraea*. The uncoupler was expected to increase TCA cycle flux and respiration and reduce α-ketoglutarate secretion. Although α-ketoglutarate secretion did halt the expected rise in respiration was not seen immediately, rather the rate of respiration fell at first before it rose to higher levels than seen previously. The observed effect was rather more complicated than expected, however the predicted cessation of α-ketoglutarate secretion and increase in respiration were both seen, however not exactly as expected. No comparable work investigating the effect of uncouplers on organic acid production has been reported, however some published work does give support to the hypothesis. Hockenhull *et al.* (1954) have demonstrated the link between limitation of pyruvate dehydrogenase and α-ketoglutarate dehydrogenase and organic acid secretion. They added inhibitors of these enzymes to a culture of *Streptomyces griseus* leading to increase secretion of pyruvate and α-ketoglutarate. Avignone Rossa *et al.* (2002) have demonstrated a link between NAD^+/NADH levels and organic acid production. They have showed that forcing the cell to make high levels of NADH by choice of carbon source leads to increased production of organic acids in *Streptomyces lividans*.

7.4 Overproduction of Erythromycin

Flux distributions in the erythromycin overproducing strain CA340 have been compared with flux distributions in the wild type organism (see chapter 5). The comparison showed that the industrial improvement of CA340 led to a strain that utilises glucose more slowly than the parental strain especially in the stationary phase. This gave a longer production phase, two mechanisms seem to have been responsible for this. Firstly the length of time glucose was available in the stationary phase was increased thereby increasing the length of the production phase. Secondly CA340 behaved as though it was glucose-limited in the exponential phase, producing erythromycin. This moved the start of the production phase to earlier in the fermentation. The lengthening of the production phase through slowing glucose uptake provided more time for higher titres of erythromycin to be produced.

Although strain CA340 has improved erythromycin production compared with the wild type it probably does not represent a good starting point for further strain improvement.

Increased yields of erythromycin have been achieved in CA340 by improving the carbon and energy efficiency through reduced glucose uptake. This allows significant improvements in titres of erythromycin. However if much larger increases in erythromycin production are to be made, the cell will have to increase its secretion of carbon, directing it to erythromycin rather than organic acids. By increasing the efficiency of strain CA340 the properties of the strain may have been made less like those required for maximum production of erythromycin.

The comparison of overproducing strains with wildtype strains in the way done above is known as reverse metabolic engineering. This work has illustrated some of the strengths and weaknesses of the approach for overproduction of secondary metabolites. Using reverse metabolic engineering it has been possible to identify a potential approach to overproducing polyketides that may be generally applicable to related organisms. This approach would work by reducing the uptake of glucose so as to increase the length of the production phase. Using this technique it may be possible to achieve modest but significant increases in production by making relatively simple genetic changes to organisms. Reverse metabolic engineering can however only identify strategies that have already been developed by classical means. For secondary metabolite production this puts restrictions on the levels of production likely to be achieved. Reverse metabolic engineering is unlikely to reveal approaches able to achieve maximum theoretical yields of secondary metabolites.

Comparison of flux distributions for wild type organisms with those required for maximal theoretical yield of erythromycin reveals radical differences. For the maximum theoretical yield calculated using elementary mode analysis (see section 4.3.6) the precursors of polyketide synthesis are supplied from three sources, valine, isoleucine and succinyl-CoA. The succinyl-CoA is synthesised using a split TCA cycle. This can only be sustained because flux through the other two routes regenerates FADH required by succinate dehydrogenase when operating in reverse. The flux through the metabolic pathways has to be tightly inter-linked in order to achieve this yield. Radically different flux distributions such as this will be hard to introduce by metabolic engineering as the flux through the different pathways will have to be inter-linked. Such distributions are also unlikely to be achieved through classical mutation and selection techniques because they are so far removed from what is seen in the wild type

and it is difficult to impose selective pressure towards increased yield. The intrinsic flexibility of *S. erythraea*'s metabolism however gives some hope that this might be possible. If it is to be achieved the best approach will probably be to make carefully chosen specific genetic changes which will allow strong selective pressure to be applied to product formation.

7.5 Final Conclusions

A metabolic network for *S. erythraea* which includes pathways relevant to stationary phase metabolism as well as growth phase metabolism has been constructed (see section 2.10). Pathways for the catabolism of biomass components such as protein and RNA have been included. Pathways for the accumulation and reutilisation of storage compounds such as trehalose and triacylglycerides have been included. Pathways for the synthesis of specialised biomass components such as arabinogalactan have also been included. The monomeric composition of the macromolecules of *S. erythraea* has also been estimated. This is the most detailed model of *S. erythraea* yet published.

The biomass composition of *S. erythraea* has been reported for the first time in chapter 3. The biomass contains protein, RNA, DNA, trehalose, arabinogalactan, lipid and a capsule possibly composed of polyglutamate. The biomass composition was measured throughout nitrogen limited batch culture and was found to vary considerably in the stationary phase as protein and RNA were broken down and trehalose, lipid and a capsule were synthesised. The formation of a capsule has not been reported in *S. erythraea* before. It is thought that the capsule might be similar to the polyglutamate capsule synthesised by *Mycobacterium tuberculosis* (see section 3.3.4.3).

The impact of including data regarding the changing biomass composition on the results obtained by flux balance analysis was also analysed in chapter 3. Inclusion of data on the biomass composition in the model was found to make a significant impact on the results achieved. The results achieved without the use of this data were not only significantly different but also were less consistent with the rest of the measured data used to calculate the results.

During the course of this project a new stationary phase product of *S. erythraea* was found. Fumaric acid was produced at low levels before glucose depletion in the stationary phase, it was reconsumed after glucose depletion (see section 4.2.1.1). This is the first time fumarate has been reported as a product of *S. erythraea* and its production profile seems similar to two other organic acids pyruvate and α-ketoglutarate.

In chapter 6 the link between oxidative phosphorylation, flux through the TCA cycle and organic acid secretion was investigated by adding an uncoupler to stationary phase cells. The rate of flux through the TCA cycle was thought to be limited by the rate of utilisation of NADH in oxidative phosphorylation leading to the secretion of α-ketoglutarate and pyruvate. The uncoupler was expected to increase respiration, increase TCA cycle flux and decrease α-ketoglutarate and pyruvate secretion. The actual effects were initially not as expected however they became more similar to the expected results after around 20 minutes. It is thought that the uncoupler concentration used interfered with glycolysis masking the expected effects in the TCA cycle until the cell started to recover. In chapter 4, calculated maintenance energy requirements for *S. erythraea* white variant wild type were found to vary greatly in the stationary phase. From this it was concluded that early in the stationary phase when glucose is in excess energy is used less efficiently than later in the stationary phase when glucose becomes limiting.

In chapters 4 and 5 it was noted that glucose uptake rate did not seem to be regulated in line with the cells requirements for carbon and energy. This observation was confirmed in chapter 6 when addition of uncoupler to a stationary phase culture of *S. erythraea* had no impact on glucose uptake despite slowing glycolysis. It is thought that the uncontrolled uptake of glucose whilst glucose is in excess during the stationary phase may be the cause of increase maintenance energy utilisation. It may also be the cause of organic acid secretion when combined with the limitation in the TCA cycle noted above.

In chapter 5 it was shown that the industrial strain CA340 has a slower uptake rate for glucose than the wild type parental strain. It also produced less organic acids and had lower maintenance energy requirements in the stationary phase than the parental strain.

It was concluded that the mutation and screening process used to produce CA340 had produced a strain which utilised carbon and energy more efficiently by reducing the uptake rate of glucose. This increased the production phase by increasing the length of time glucose was available in the stationary phase. It may also have caused partial glucose limitation in the growth phase leading to erythromycin production in this phase. It is thought that it may be possible to introduce this secondary metabolite overproduction strategy to other related organisms quite simply. This should be done by reducing the expression of glucose uptake systems or reducing glucose uptake using fed batch systems.

Comparison of flux split ratios at key points in the metabolic network of *S. erythraea* at different stages in the fermentation revealed that flux split in the network is not rigidly controlled (see section 4.3.4.5). This was confirmed by the uncoupler experiment where rapid changes in the flux split ratio at two key branch points in the network were seen on addition of the uncoupler (see section 6.3.5). It is thought that inbuilt metabolic flexibility of the *S. erythraea* metabolic network should facilitate metabolic engineering of this organism.

The flux distributions leading to the maximum production of erythromycin on glucose and protein or RNA were determined by elementary modes analysis (see section 4.3.6). The flux distributions with highest yield were similar to each other but quite different from those seen in the wild type organism. The distributions with the maximum yields utilised a split TCA cycle and derived monomers for polyketide synthesis from isoleucine, valine and succinyl-CoA in a complex distribution.

Applying flux balance analysis to *S. erythraea* in a way which takes into account the organism's natural biochemistry has yielded a number of important insights into the metabolism of the organism. This in turn has lead to the identification of strategies for the overproduction of erythromycin, which may be applicable to other related organisms producing secondary metabolites.

7.6 Future Work

A number of areas that have been investigated in this project warrant further research. The suggested link between lowered glucose uptake rates and higher levels of erythromycin production is one such area. Ultimately it would be desirable to genetically engineer the organism to have lower glucose uptake, however a number of quicker and simpler approaches could be used for small-scale preliminary investigations. Glucose uptake can be controlled in chemostats or in fed batch fermentations. Alternatively slow release carbon sources such as starch could be used. Another approach would be to use a glucose analogue which would compete with glucose for uptake but would not be metabolisable. Woodrow *et al.* (2000) tested a number of such compounds for competitive inhibition of glucose uptake in a *Plasmodium* species. All these options have advantages and disadvantages that would have to be weighted before a strategy is adopted. Making a genetic change to the organism would take longer than the other strategies but would give a strain that can be grown in batch culture as though limited for carbon without the use of inhibitors or unusual carbon sources. Use of inhibitors is simple but would be expensive on anything but the smallest scale. Use of a chemostat allows accurate and variable control of glucose up take, however it puts strains under great selective pressure and has not found much use at the industrial scale. Use of slow release carbon sources has been practised in industrial fermentation for secondary metabolite production (Stanbury *et al.* 1995). Carbon sources such as starch are cheap and simple to use. This method however is effectively lowering the concentration of glucose available rather than the cells ability to take it up and this may alter the cell's response.

The work done on investigating the restrictions on flux through the TCA cycle using uncouplers could be extended in a couple of areas. Firstly investigation of a wider range of concentrations of uncoupler, particularly lower concentrations might be interesting. In the work presented here there is some evidence to support the hypothesis that the coupling of NADH production and oxidative phosphorylation is restricting TCA cycle flux. To a large extent however the effect of adding the uncoupler on the TCA cycle is masked because flux through glycolysis is restricted when the uncoupler is added. Lower concentrations of uncoupler might not have such an adverse effect on

glycolysis allowing the effect on the TCA cycle to be seen more clearly. Hockenhull *et al.* (1954) found that addition of uncoupler to *Streptomyces griseus* increased respiration. It may be that with lower concentrations of uncoupler the same response will be seen in *S. erythraea.* It would also be useful to confirm where the excess glucose taken into the cell is going whilst glycolysis is slowed. The evidence presented here suggests it is stored as a sugar, possibly trehalose. This could be tested experimentally by measuring the trehalose levels before and after the uncoupler is added.

The discovery of a capsule synthesised in the stationary phase in *S. erythraea* is another area where further research could be performed. The evidence points to the capsule being composed of polyglutamate, this should be tested experimentally. Methods for purifying polyglutamate exist (Ashiuchi *et al.* 1999) however they are long and complex and not well suited to handling large numbers of samples. It may not be necessary to isolate the polyglutamate in order to identify and quantify it. Polyglutamate is a polypeptide but rather than being α-linked like normal proteins it is linked via the carboxyl group on the γ carbon. This makes it resistant to normal peptidase enzymes. Enzymes that depolymerise polyglutamate do exist (Abe *et al.* 1997) however they are not commercially available. It is not therefore possible to specifically degrade it. It is however possible to hydrolyse it using the conditions applied for the hydrolysis of normal proteins (Ashiuchi *et al.* 1999). The cell wall components of biomass can be isolated from the soluble protein pool of the cell using the extraction method of Mousdale (1998). Quantification of all amino acids in the cell wall would allow the total amount of glutamate in the cell wall to be determined. Comparison of this with the levels of the other amino acids in the cell wall over the course of a fermentation should reveal the synthesis of a polyglutamate capsule. An alternative method would be to knock out the genes responsible for polyglutamate synthesis and to check for capsule formation in the stationary phase. It would also be interesting to see what effect this has on the stationary phase behaviour of the organism. If it reduces wall growth and viscosity this may well have a beneficial impact on the fermentation. It will also be interesting to note where the excess nitrogen is channelled in the absence of polyglutamate formation, whether it drives further growth or merely accumulates in the broth.

A theoretical analysis of the metabolism of *S. erythraea* suggested that the maximum yield of erythromycin would be achieved by operating a split TCA cycle. It would be interesting to determine whether *S. erythraea* can grow with a split TCA cycle. In order to do this appropriate conditions would have to be designed. One approach would be to try to supply the cells requirement for NADH from the pentose phosphate pathway so that no TCA cycle activity is required. If the cell is to grow TCA cycle intermediates will still be required and in the absence of normal TCA cycle activity these could only be supplied by operating a split TCA cycle. This strategy requires the flux through the PPP to be increased as much as possible. One way to do this is to increase the growth associated NADPH requirements as much as possible. The use of nitrate as nitrogen source maximises the NADPH requirements for nitrogen assimilation. It is also possible to try to reduce flux through enzymes that regenerate NADPH. Avignone Rossa et al. (2002) achieved this by using gluconate as carbon source in place of glucose. This reduces flux through glucose-6-phosphate dehydrogenase, which produces NADPH because gluconate enters the PPP directly. Its catabolism still however still utilises 6-phosphoguluconate dehydrogenase which produces NADH. Therefore the amount of NADH produced in the PPP will go up as greater recycling through the PPP will be required to generate the same amount of NADPH. It ought to be noted however that in the case of Avignone Rossa *et al.* (2002) this merely led to increased organic acid production. Some of the flux distributions determined in this work already show rather low TCA cycle activity. By combining gluconate as carbon source with nitrate as nitrogen source it might be possible to halt normal TCA cycle activity causing a split TCA cycle.

In chapter 4.3.4.1.2 it was found that if the TCA cycle were to supply the NADH requirements of oxidative phosphorylation during the stationary phase this would lead to overproduction of NADPH. Speculation was made as to how the cell avoids this problem however with the methods available it was not possible to test the different possibilities envisaged. A pathway formed by cycling pyruvate through malic enzyme and phosphoenolpyruvate carboxykinase having a net transhydrogenase effect was proposed (see section 4.3.4.1.2). Another alternative involving anaplerotic pathways found in *E. coli* by Fischer and Sauer (2003) could also alleviate the problem. It would be interesting to elucidate what actually happens in *S. erythraea* during the stationary phase. This could be investigated by feeding labelled glucose and analysing the amino

acids produced by mass spectrometry as was done for *E. coli* by Fischer and Sauer (2003). The flux split at key branch points can then be determined from the ratios of different labeling patterns in the amino acids. This combined with flux balance analysis can provide flux distributions for the pathways in question. However it is the stationary phase that is of greatest interest and during this phase amino acid production does not occur. It may however be possible to perform the same analysis during the stationary phase using other products such as the secreted organic acids. The elucidation of this pathway could have large implications for strategies to overproduce erythromycin because succinyl-CoA, a TCA cycle intermediate is one of the precursors of erythromycin production. Directing the flux to this compound could prove key to overproducing erythromycin therefore answering the question of whether a classic TCA cycle or a less well known cycle is operating in stationary phase *S. erythraea* could have a big impact on the approach taken to overproduce erythromycin.

8 REFERENCES

Abe, K., Ito, Y., Ohmachi, T. and Asada, Y. (1997)
Purification and properties of two isozymes of γ-glutamyltranspeptidase from *Bacillus subtilis* TAM-4.
Bioscience, Biotechnology and Biochemistry 61, 1621-1625

Ahmad, S., Selvapandiyan, A. and Bhatnagar, R. (1999)
A protein-based phylogenetic tree for Gram-positive bacteria derived from *hrcA*, a unique heat-shock regulatory gene
International Journal of Systematic Bacteriology 49, 1387-1394

Ahmad, S., Selvapandiyan, A. and Bhatnagar, R. (2000)
Phylogenetic analysis of Gram-positive bacteria based on *grpE*, encoded by the *dnaK* operon
International Journal of Systematic and Evolutionary Microbiology 50, 1761-1766

Aisaka, K. and Masuda, T. (1995)
Production of trehalose phosphorylase by Catellatospora ferruginea
FEMS Microbiology Letters 131, 47-51

Alvarado, A. and Flores, M. (2003)
Characterization and regulation of $NADP^+$-isocitrate dehydrogenase from *Saccharopolyspora erythraea*

Ashiuchi, M., Soda, K. and Misono, H. (1999)
A poly-g-glutamate synthetic system of *Bacillus subtilis* IFO 3336: Gene cloning and biochemical analysis of poly-g-glutamate produced by *Escherichia coli* clone cells
Biochemical and Biophysical Research Communications 263, 6-12

Avignone Rossa, C., White, J., Kuiper, A., Postma, P., Bibb, M. and Teixeira de Mattos, M. (2002)
Carbon flux distribution in antibiotic-producing chemostat cultures of *Streptomyces lividans*
Metabolic Engineering 4,138-150

Bailey, J., Sburlati, A., Hatzimanikatis, V., Lee, K., Renner, W. and Tsai, P. (1996)
Inverse metabolic engineering: a strategy for directed genetic engineering of useful phenotypes
Biotechnology and Bioengineering 52, 109-121

Bailey, J. (1999)
Lessons from metabolic engineering for functional genomics and drug discovery
Nature Biotechnology 17, 616-618

Barona-Gomez, F., Lautru, S. and Challis, G. (2003)
Secondary metabolomics of *Streptomyces coelicolor*: linking genes with metabolites.
Abstract of a presentation, *Streptomyces* Dissemination Meeting, University of Surrey

Bate, N., Butler, A., Gandecha, A. and Cundliffe, E. (1999)
Multiple regulatory genes in the tylosin biosynthesis cluster of *Streptomyces fradiae*
Chemistry and Biology 6, 617-624

Bedford, D., Schweizer, E., Hopwood, D. and Khosla, C. (1995)
Expression of a functional fungal polyketide synthase in the bacterium *Streptomyces coelicolor* A3(2)
Journal of Bacteriology 177, 4544-4548

Benthin, S., Nielsen, J. and Villadsen, J. (1991)
A simple and reliable method for the determination of cellular RNA content
Biotechnology Techniques 5, 39-42

Bentley, S., Chater, K., Cerdeño-Tárraga, A., Challis, G., Thomson, N., James, K., Harris, D., Quail, M., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'neil, S., Rabbinowitsch, E., Rajandream, M., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B., Parkhill, J. and Hopwood, D. (2002)
Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2)
Nature 417, 141-147

Bermúdez, O., Padilla, P., Huitrón, C. and Flores, M. (1998)
Influence of carbon and nitrogen source on synthesis of $NADP^+$ -isocitrate dehydrogenase, methylmalonyl-coenzyme A mutase, and methylmalonyl-coenzyme A decarboxylase in *Saccharopolyspora erythraea* CA340
FEMS Microbiology Letters 164, 77-82

Bertram, R., Schlicht, M., Mahr, K., Nothaft, H., Saier, M. and Titgemeyer, F. (2004)
In silico and transcriptional analysis of carbohydrate uptake systems of *Streptomyces coelicolor* A3(2)
Journal of Bacteriology 186, 1362-1373

Bowen, T., Stackebrandt, E., Dorsch, M. and Embley, T. (1989)
The phylogeny of *Amycolata autotrophica*, *Kibdelosporangium aridum* and *Saccharothrix australiensis*
Journal of General Microbiology 135, 2529-2536

Bradford, M. (1976)
A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding
Analytical Biochemistry 72, 248-254

Braña, A., Méndez, C., Díaz, L., Manxanal, M. and Hardisson, C. (1986)
Glycogen and trehalose accumulation during colony development in *Streptomyces antibioticus*
Journal of General Microbiology 132, 1319-1326

Bruton, C., Plaskitt, K. and Chater, K. (1995)
Tissue-specific glycogen branching isoenzymes in a multicellular prokaryote, *Streptomyces coelicolor* A3(2)
Molecular Microbiology 18, 89-99

Bushell, M., Smith, J. and Lynch, H. (1997)
A physiological model for the control of erythromycin production in batch and cyclic fed batch culture
Microbiology 143, 475-480

Butler, M., Bruheim, P., Jovetic, S., Marinelli, F., Postma, P. and Bibb, M. (2002)
Engineering of primary carbon metabolism for improved antibiotic production in *Streptomyces lividans*
Applied and Environmental Microbiology 68, 4731-4739

Cain, D., Walsh, C. and Khosla, C. (1998)
Harnessing the biosynthetic code: Combinations, permutations and mutations
Science 282, 63-68

Cao, N., Du, J., Gong, C. and Tsao, G. (1996)
Simultaneous production and recovery of fumaric acid from immobilized Rhizopus oryzae with a rotary biofilm contactor and an adsorption column.
Applied and Environmental Microbiology 62, 2926-2931

Carreras, C. and Santi, D, 1998
Engineering of modular polyketide synthases to produce novel polyketides
Current Opinion in Biotechnology 9, 403-411

Cataldo, D., Haroon, M., Schrades, L. and Youngs, V. (1975)
Rapid colorimetric determination of nitrate in plant tissue by nitration of salicylic acid.
Communications in Soil Science and Plant Analysis 6, 71-80.

Celmer, W. (1965 a)
Macrolide stereochemistry II. Configurational assignments at certain centres in various macrolide antibiotics
Journal of the American Chemical Society 87, 1799-1800

Celmer, W. (1965 b)
Macrolide stereochemistry. III. A Configurational model for macrolide antibiotics
Journal of the American Chemical Society 87, 1801-1802

Champney, W., Chittum, H. and Tober, C. (2003)
A 50S ribosomal subunit precursor particle is a substrate for the ErmC methyltransferase in *Staphylococcus aureus* cells
Current Microbiology 46, 453-460

Chapman, A., Fall, A. and Atkinson, D. (1971)
Adenylate energy charge in *Escherichia coli* during growth and starvation
Journal of Bacteriology 108, 1072-1086

Chatterji, D. and Ojha, A. (2001)
Revisiting the stringent response, ppGpp and starvation signalling
Current Opinion in Microbiology 4, 160-165

Christensen, B. and Nielsen, J. (1999)
Metabolic network analysis
Advances in Biochemical Engineering/Biotechnology 66, 209-231

Christiansen, T., Christensen, B. and Nielsen, J. (2001)
Metabolic network analysis of *Bacillus clausii* on minimal and semirich medium using ^{13}C-labaled glucose.
Metabolic Engineering 4, 159-169

Cole, S., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S., Eiglmeier, K., Gas, S., Barry, C., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver K, Osborne J, Quail M, Rajandream, M, Rogers J, Rutter S, Seeger K, Skelton J, Squares, S., Squares, R., Sulston, J., Taylor, K., Whitehead, S. and Barrell, B. (1998)
Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence.
Nature 393, 537-544

Cortés, J., Wiesmann, K., Roberts, G. Brown, M., Staunton, J. and Leadlay, P. (1995)
Repositioning of a domain in a modular polyketide synthase to promote specific chain cleavage
Science 268, 1487-1489

Cortés, J., Velasco, J., Foster, G., Blackaby, A., Rudd, B. and Wilkinson, B. (2002)
Identification and cloning of a type III polyketide synthase required for diffusable pigment biosynthesis in *Saccharopolyspora erythraea*
Molecular Microbiology 44, 1213-1224

Crick, D., Mahapatra, S. and Brennan, P. (2001)
Biosynthesis of the arabinogalactan-peptidoglycan complex of *Mycobacterium tuberculosis*
Glycobiology 11, 107R-118R

Daae, E. (1999)
Mathematical modelling of biochemical pathways
PhD thesis, University of London

Daae, E. and Ison, A. (1999)
Classification and sensitivity analysis of a proposed primary metabolic reaction network for *Streptomyces lividans*.
Metabolic Engineering 1, 153-165

Davidson, A. (1992)
Quantitative microbial physiology of *Streptomyces* coelicolor A3(2)
PhD thesis, University of Glasgow, UK

Davies, J., Baganz, F., Ison, A. and Lye, G. (2000).
Studies on the interaction of fermentation and microfiltration operations: erythromycin recovery from *Saccharopolyspora erythraea* fermentation broths
Biotechnology and Bioengineering 69, 429-439

Dekleva, M. and Strohl, W. (1988a)
Biosynthesis of ε-rhodomycinone from glucose by *Streptomyces* C5 and comparison with intermediary metabolism of other polyketide-producing streptomycetes
Canadian Journal of Microbiology 34, 1235-1240

Dekleva, M. and Strohl, W. (1988b)
Activity of phosphoenolpyruvate carboxylase of an anthracycline-producing streptomycete
Canadian Journal of Microbiology 34, 1241-1246

Donadio, S., Staver, M. and Katz, L. (1996)
Erythromycin production in *Saccharopolyspora erythraea* does not require a functional propionyl-CoA carboxylase
Molecular Microbiology 19, 977-984

Doskočil, J., Hošťálek, Z., Kašparová, J., Zajíček, J. and Herold, M. (1959)
Development of *Streptomyces aureofaciens* in submerged culture
Journal of Biochemical and Microbiological Technology and Engineering 1, 261-271

Du, L., Sánchez, C. and Shen, B. (2001)
Hybrid peptide-polyketide natural products: biosynthesis and prospects towards engineering novel molecules
Metabolic Engineering 3, 78-95

Duboc, Ph., Schill, N., Menoud, L., van Gulik, W. and von Stockar, U. (1995)
Measurements of sulfur, phosphorous and other ions in microbial biomass: influence on correct determination of elemental composition and degree of reduction
Journal of Biotechnology 43, 145-158

Edwards, J. and Palsson, B. (2000)
The *Escherichia coli* MG1655 *in silico* metabolic genotype: Its definition, characteristics, and capabilities
Proceedings of the National Academy of Science (USA) 97, 5528-5533

Elbein, A., Pan, Y., Pastuszak, I. and Carroll, D. (2003)
New insights on trehalose: a multifunctional molecule
Glycobiology 13, 17R-27R

Elibol, M. and Mavituna, F. (1999)
A kinetic model for actinorhodin production by *Streptomyces coelicolor* A3(2)
Process Biochemistry 34, 625-631

Embley, M., Smida, J. and Stackebrandt, E. (1988a)

The phylogeny of mycolate-less wall chemotype IV actinomycetes and description of *Pseudonocardiaceae* fam. nov.
Systematic Applied Microbiology Volume 11, 44-52

Embley, M., Smida, J. and Stackebrandt, E. (1988b)
Reverse transcriptase sequencing of 16S ribosomal RNA from *Faenia rectivirgula*, *Pseudonocardia thermophila* and *Saccharopolyspora hirsuta*, three wall IV actinomycetes which lack mycolic acids
Journal of General Microbiology 134, 961-966

Embley, M. and Stackebrandt, E. (1994)
The molecular phylogeny and systematics of the actinomycetes
Annual Review of Microbiology 48, 257-289

Fell, D. (2001)
Beyond genomics
Trends in Genetics 17, 680-682

Fischer, E. and Sauer, U. (2003)
A novel metabolic cycle catalyses glucose oxidation and anaplerosis in hungry *Escherichia coli*
Journal of Biological Chemistry 278, 46446-46451

Flores, N., Xiao, J., Berry, A., Bolivar, F. and Valle, F. (1996)
Pathway engineering for the production of aromatic compounds in *Escherichia coli*
Nature Biotechnology 14, 620-623

Fothergill, and Guest, (1977)
Catabolism of L-Lysine by Pseudomonas aeruginosa
Journal of General Microbiology 99, 139-155

Funa, N., Ohnishi, Y., Fujii, I., Shibuya, M., Ebizuka, Y. and Horinouchi, S. (1999)
A new pathway for polyketide synthesis in microorganisms
Nature 400, 897-899

Gaisser, S., Bhöm, G., Doumith, M., Cortés, J. and Leadlay, P (1997)
Analysis of seven genes from the *eryAI- eryK* region of the erythromycin biosynthetic gene cluster in *Saccharopolyspora erythraea*
Molecular and General Genetics 258, 78-88

Gaisser, S., Bhöm, G., Doumith, M., Raynal, M., Dhillon, N., Cortés, J. and Leadlay, P (1998)
Analysis of *eryBI*, *eryBIII* and *eryBVII* from the erythromycin biosynthetic gene cluster in *Saccharopolyspora erythraea*
Molecular and General Genetics 258, 78-88

Gaisser, S., Reather, J., Wirtz, G., Kellenberger, L., Staunton, J. and Leadlay, P. (2000)
A defined system for hybrid macrolide biosynthesis in *Saccharopolyspora erythraea*
Molecular Microbiology 36, 391-401

Gamian, A., Mordarska, H., Ekiel, I., Ulrich, J., Szponar, B. and Defaye, J. (1996)
Structural studies of the major glycolipid from *Saccharopolyspora* genus
Carbohydrate Research 296, 55-67

Gerhardt, P., Murray, R., Costilow, R., Nester, E., Wood, W., Krieg, N. and Briggs Phillips, G. (1981)
Manual of methods for general bacteriology
American Society for Microbiology

Gershater, C. (2000)
Cambridge Bioprocess Management Ltd
Personal communication

Goodfellow, M., Lacey, J., Athalye, M., Embley, T. and Bowen, T. (1989)
Saccharopolyspora gregorii and *Saccharopolyspora hordei*: two new actinomycete species from fodder
Journal of General Microbiology 135, 2125-2139

Gonzalez, B., François, J. and Renaud, M. (1997)
A rapid and reliable method for metabolite extraction in yeast using boiling buffered ethanol
Yeast 13, 1347-1356

Groussac, E., Ortiz, M. and François, J. (2000)
Improved protocols for quantitative determination of metabolites from biological samples using high performance ion-exchange chromatography with conductiometric and pulsed amperometric detection
Enzyme and Microbial Technology 26, 715-723

Hajjaj, H., Blanc, P., Goma, G. and François, J. (1998)
Sampling techniques and comparative extraction procedures for quantitative determination of intra- and extracellular metabolites in filamentous fungi
Fems Microbiology letters 164, 195-200

Han, L. and Reynolds, K. (1997)
A novel alternate anaplerotic pathway to the glyoxylate cycle in streptomycetes
Journal of Bacteriology 179, 5157-5164

Harth, G. and Horwitz, M. (1999)
An Inhibitor of Exported *Mycobacterium tuberculosis* Glutamine Synthetase Selectively Blocks the Growth of Pathogenic *Mycobacteria* in Axenic Culture and in Human Monocytes: Extracellular Proteins as Potential Novel Drug Targets
Journal of Experimental Medicine 189, 1425-1436

Harth, G., Zamecnik, P., Tang, J., Tabatadze, D. and Horwitz, M. (2000)
Treatment of *Mycobacterium tuberculosis* with antisense oligonucleotides to glutamine synthetase mRNA inhibits glutamine synthase activity, formation of the poly-L-glutamate/glutamine cell wall structure, and bacterial replication
Proceedings of the National Academy of Science 97, 418-423

Herbert, D., Phipps, P. and Strange, R. (1971)
Chemical analysis of microbial cells
Methods in Microbiology 5b, 209-344

Hesketh, A., Chandra, G., Shaw, A., Rowland, J., Kell, D., Bibb, M. and Chater, K. (2002)
Primary and secondary metabolism, and post-translational protein modifications, as portrayed by proteomic analysis of *Streptomyces coelicolor.*
Molecular Microbiology 46, 917-932

Heydarian, S., Ison, A. and Mirjalili, N. (1998).
A rapid and simplified extraction method of erythromycin from fermentation broth with bond elut C18 cartridge for analysis by HPLC
Biotechnology Techniques 12, 155-158

Hockenhull, D., Fantes, K., Herbert, M. and Whitehead, B. (1954)
Glucose utilization by *Streptomyces griseus*
Journal of General Microbiology 10, 353-370

Hodgson, D. (2000)
Primary metabolism and its control in *streptomyces*: a most unusual group of bacteria
Advances in Microbial Physiology 42, 49-238

Hsieh, Y. and Kolattukudy, P. (1994)
Inhibition of erythromycin synthesis by disruption of malonyl-coenzyme A decarboxylase gene *eryM* in *Saccharopolyspora erythraea*
Journal of Bacteriology 176, 714-724

Hu, Z., Hopwood, D. and Hutchinson, R. (2003)
Enhanced heterologous polyketide production in *Streptomyces* by exploiting plasmid co-integration
Journal of Industrial Microbiology and Biotechnology 30, 516-522

Huang, J., Lih, C., Pan, K. and Cohen, S. (2001)
Global analysis of growth phase responsive gene expression and regulation of antibiotic biosynthetic pathways in *Streptomyces coelicolor* using DNA microarrays.
Genes and Development 15, 3183-3192

Huynen, M., Snel, B., Lathe, W III. and Bork, P. (2000)
Exploitation of gene context
Current Opinion in Structural Biology 10, 366-370

Ideker, T., Thorsson, V., Ranish, J., Christmas, R., Buhler, J., Eng, J., Bumgarner, R., Goodlett, D., Aebersold, R. and Hood, L. (2001)
Integrated genomic and proteomic analyses of a systematically perturbed metabolic network
Science 292, 929-934

Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M. and Omura, S. (2003)

Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*.
Nature Biotechnology 21, 526-531

Inbar, L. and Lapidot, S. (1991)
^{13}C nuclear magnetic resonance and gas chromatography-mass spectrometry studies of carbon metabolism in the actinomycin D producer *Streptomyces parv*ulus by use of ^{13}C-labeled precursors
Journal of Bacteriology 173, 7790-7801

Ito, M., Ohnishi, Y., Itoh, S. and Nishimura, M. (1982)
Carbonyl Cyanide-m-chlorophenyl hydrazone resistant *Escherichia coli* mutant that exhibits a temperature sensitive Unc phenotype.
Journal of Bacteriology 153, 310-315.

Ives, P. and Bushell, M. (1997)
Manipulation of the physiology of clavulanic acid production in *Streptomyces clavuligerus*
Microbiology 143, 3573-3579

Izumikawa, M., Shipley, P., Hopke, J., O'Hare, T., Xiang, L., Noel, J. and Moore, B. (2003)
Expression and characterization of the type III polyketide synthase 1,3,6,8-tetrahydroxynaphthalene synthase from *Streptomyces coelicolor* A3(2)
Journal of Industrial Microbiology and Biotechnology 30, 510-515

Jeong, H., Tombor, B., Albert, R., Oltvai, Z. and Barabási, A. (2000)
The large scale organisation of metabolic networks
Nature 407, 651-654

Jonsbu, E., Christensen, B. and Nielsen, J. (2001)
Changes of *in vivo* fluxes though central metabolic pathways during the production of nystatin by *Streptomyces noursei* in batch culture
Applied Microbiology and Biotechnology 56, 93-100

Jørgensen, H., Nielson, J., Møllgaard, H. and Villadsen, J. (1995)
Metabolic flux distributions in *Penicillium chrysogenum* during fed batch cultivations
Biotechnology and Bioengineering 46, 117-131

Kao, C., Luo, G., Katz, L., Cane, D. and Khosla, C. (1994)
Engineered biosynthesis of a triketide lactone from an incomplete modular polyketide synthase
Journal of the American chemical Society 116, 11612-11613

Kalinowski, J., Bathe, B., Bartels, D., Bischoff, N., Bott, M., Burkovski, A., Dusch, N., Eggeling, L., Eikmanns, BJ., Gaigalat, L., Goesmann, A., Hartmann, M., Huthmacher, K., Kramer, R., Linke, B., McHardy, A., Meyer, F., Mockel, B., Pfefferle, W., Puhler, A., Rey, D., Ruckert, C., Rupp, O., Sahm, H., Wendisch, V., Wiegrabe, I. and Tauch, A. (2003)

The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins.
Journal of Biotechnology 104, 5-25

Kanehisa, M. and Goto, S. (2000)
KEGG: Kyoto Encyclopedia of Genes and Genomes
Nucleic Acids Research 28, 27-30

Kannan, L. and Rehacek, Z. (1970)
Metabolism of *Streptomyces antibioticus*: variation of organic and amino acids
Indian Journal of Biochemistry 7, 137-138

Karamanou, S., Vrontou, E., Sianidis, G., Baud, C., Roos, T., Kuhn, A., Politou, A. and Economou, A. (1999)
A molecular switch in SecA protein couples ATP hydrolysis to protein translocation
Molecular Microbiology 34, 1133-1145

Karandikar, A., Sharples, G. and Hobbs, G. (1997)
Differentiation of *Streptomyces coelicolor* A3(2) under nitrate-limited conditions
Microbiology 143, 3581-3590

Karp, P., Krummenacker, M., Paley, S. and Wagg, J. (1999)
Integrated pathway-genome databases and their role in drug discovery
Trends in Biotechnology 17, 275-281

Katz, L. (1997)
Manipulation of modular polyketide synthases
Chemical Reviews 97, 2557-2575

Kealey, J, Liu, L., Santi, D., Betlach, M. and Barr, P. (1998)
Production of a polyketide natural product in nonpolyketide producing prokaryotic and eukaryotic hosts
Proceedings of the National Academy of Science (USA) 95, 505-509

Khosla, C. (1997)
Harnessing the biosynthetic potential of modular polyketide synthases
Chemical Reviews 97, 2577-2590

Khosla, C. and Bailey, J. (1988)
Heterologous expression of a bacterial hemoglobin improves the growth properties of recombinant *Escherichia coli*
Nature 331, 633-635

King, R. (1997)
A structured mathematical model for a class of organisms: I. Development of a model for *Streptomyces tendae* and application of model-based control
Journal of Biotechnology 52, 219-234

King, R. and Büdenbender, Ch. (1997)

A structured mathematical model for a class of organisms: II. Application of the model to other strains
Journal of Biotechnology 52, 235-244

Kirk, S., Avignone-Rossa, C. and Bushell, M. (2000)
Growth limiting substrate affects antibiotic production and associated metabolic fluxes in *Streptomyces clavuligerus*
Biotechnology Letters 22, 1803-1809

Kirschning, A., Bechthold, A. and Rohr, J. (1997)
Chemical and biochemical aspects of deoxysugars and deoxysugar oligosaccharides
Topics in Current Chemistry 188, 1-84

Klamt, S., Stelling, J., Ginkel, M. and Gilles, E. (2003)
FluxAnalyzer: exploring structure, pathways, and flux distributions in metabolic networks on interactive flux maps
Bioinformatics 19, 261-269

Koebmann, B., Westerhoff, H., Snoep, J., Nilsson, D. and Jensen, P. (2002)
The glycolytic flux in *Escherichia coli* is controlled by the demand for ATP
Journal of Bacteriology 184, 3909-3916

Kuznetsov, V., Rodionova, E., Yangulova, I. and Dmitrieva, S. (1977)
Review of the systematic position of *Actinomyces erythraeus* Waksman et Curtis 1916 and its transfer to the genus *Proactinomyces*
Mikrobiologiya 46, 137-142

Labeda, D. (1987)
Transfer of the type strain of *Streptomyces erythraeus* (Waksman 1923) Waksman and Henrici 1948 to the genus *Saccharopolyspora* Lacey and Goodfellow 1975 as *Saccharopolyspora erythraea* sp. nov., and designation of a neotype strain for *Streptomyces erythraeus*
International Journal of Systematic Bacteriology 37, 19-22

Lal, R., Khanna, R., Kaur, H., Khanna, M., Dhingra, N., Lal, S., Gartemann, K., Eichenlaub, R. and Ghosh, P. (1996)
Engineering antibiotic producers to overcome the limitations of classical strain improvement programs
Critical Reviews in Microbiology 22, 201-255

Lal, R., Kumari, R., Kaur, H., Khanna, R., Dhingra, N. and Tuteja, D. (2000)
Regulation and manipulation of the gene clusters encoding type-1 PKSs
Trends in Biotechnology Volume 18, 264-274

Lange, H. and Heijnen, J. (2001)
Statistical reconciliation of the elemental and molecular biomass composition of Saccharomyces cerevisiae
Biotechnology and Bioengineering 75, 334-344

Leadlay, P., Staunton, J., Aparicio, J., Bevitt, D., Caffrey, P., Cortés, J., Marsden, A. and Roberts, G. (1993)
The erythromycin-producing polyketide synthase
Biochemical Society Transactions 21, 218-222

Lechevalier, M., De Bievre. and Lechevalier, H. (1977)
Chemotaxonomy of aerobic actinomycetes: phospholipid composition
Biochemical Systematics and Ecology 5, 249-260

Liao, X., Vining, L. and Doull, J. (1995)
Physiological control of trophophase-idiophase separation in streptomycete cultures producing secondary metabolites
Canadian Journal of Microbiology 41, 309-315

Liu, H. and Thorson, J (1994)
Pathways and mechanisms in the biogenesis of novel deoxysugars by bacteria
Annual Review of Microbiology 48, 223-256

Liu, H. and Reynolds, K. (2001)
Precursor supply for polyketide biosynthesis: the role of crotonyl-CoA reductase
Metabolic Engineering 3, 40-48,

Liu, L., Thamchaipenet, A., Fu, H., Betlach, M. and Ashley, G. (1997)
Biosynthesis of 2-nor-6-deoxyerythronolide B by rationally designed domain substitutions
Journal of the American Chemical Society 119, 10553-10554

Lomovskaya, N., Otten, S., Doi-katayama, Y., Fonstein, L., Liu, X., Takatsu, T., Inventi-solari, A., Filippini, S., Torti, F., Colombo, A. and Hutchinson, R. (1999)
Doxorubicin overproduction in *Streptomyces peucetius*: cloning and characterisation of the *dnrU* ketoreductase and *dnrV* genes and the *doxA* cytochrome P-450 hydroxylase gene
Journal of Bacteriology 181, 305-318

Luchsinger, W. and Cornesky, R. (1962)
Reducing power by the dinitrosalicylic acid method
Analytical Biochemistry 4, 346-347

Lynch, H. and Bushel, M. (1995)
The physiology of erythromycin biosynthesis in cyclic fed batch culture
Microbiology 141, 3105-3111

Madden, T., Ward, J. and Ison, A. (1996)
Organic acid excretion by *Streptomyces lividans TK24* during growth on defined carbon and nitrogen sources.
Microbiology 142, 3181-3185

Madron, F., Veverka, V. and Vanecek, V. (1977)
Statistical analysis of material balance of a chemical reactor
AIChE Journal 23, 482-486

Marsden, A., Wilkinson, B., Cortés, J., Dunster, N., Staunton, J. and Leadlay, P. (1998)
Engineering broader specificity into an antibiotic-producing polyketide synthase
Science 279, 199-202

Martin, C., Schneider, D., Bruton, C., Charter, K, and Hardisson C. (1997)
A *glgC* gene essential only for the first of two spatially distinct phases of glycogen synthesis in *Streptomyces coelicolor* A3(2)
Journal of Bacteriology 179, 7784-7789

Marx, A., Striegel, K., de Graaf, A., Sahm, H. and Eggeling, L. (1997)
Response of the central metabolism of *Corynebacterium glutamicum* to different flux burdens
Biotechnology and Bioengineering 56, 168-180

McDaniel, R., Kao, C., Fu, H., Hevezi, P., Gustafsson, C., Betlach, M., Ashley, G., Cane, D. and Khosla, C. (1997)
Gain-of-function mutagenesis of a modular polyketide synthase
Journal of the American Chemical Society 119, 4309-4310

McDaniel, R., Thamchaipenet, A., Gustafsson, C., Fu, H., Betlach, M., Betlach, M. and Ashley, G. (1999)
Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products
Proceedings of the National Academy of Science (USA) 96, 1846-1851

McDaniel, R., Licari, P. and Khosla, C. (2001)
Process development and metabolic engineering for the overproduction of natural and unnatural polyketides
Advances in Biochemical Engineering and Biotechnology 73, 31-52

McDermott, J., Lethbridge, G. and Bushell, M. (1993)
Estimation of the kinetic constants and elucidation of trends in growth and erythromycin production in batch and continuous cultures of *Saccharopolyspora erythraea* using curve-fitting techniques
Enzyme and Microbial Technology 15, 657-663

McGuire, J., Bunch, R., Anderson, R., Boaz, H., Flynn, E., Powell, H. and Smith, J. (1952)
Ilotycin, a new antibiotic
Antibiot. Chemother. Volume 2, 281-283

McKay, D., Davies, M. and Berry, S. (2001)
Bacterial genome sequenced
Trends in Biotechnology 19, 246

Mesnage, S., Tosi-Couture, E., Gounon, P., Mock, M. and Fouet, A. (1998)
The capsule and S-layer: toe independent and yet compatible macromolecular structures in *Bacillus anthracis*
Journal of Bacteriology 180, 52-58

Miguélez, E., Fernández, M. and Hardisson, C. (1997)
Nitrogen starvation-induced glycogen synthesis depends on the developmental stage of *Streptomyces antibioticus* mycelium
FEMS Microbiology Letters 153, 57-62

Minas, W., Brünker, P., Kallio, P. and Bailey, J. (1998)
Improved erythromycin production in a genetically engineered industrial strain of *Saccharopolyspora erythraea*
Biotechnology Progress 14, 561-566

Mira de Orduña, R. and Theobald, U. (2000)
Intracellular glucose 6-phosphate content in *Streptomyces coelicolor* upon environmental changes in defined medium
Journal of Biotechnology 77, 209-218

Molenaar, D., van der Rest, M., Drysch, A. and Yücel, R.
Functions of the membrane-associated and cytoplasmic malate dehydrogenases in the citric acid cycle of *Corynebacterium glutamicum*
Journal of Bacteriology 182, 6884-6891

Mousdale, D. (1998)
The analytical chemistry of microbial cultures
in Fermentation Microbiology and Biotechnology
Edited by Mansi E. and Bryce C.
Published by Taylor and Frances Ltd

Murli, S., Kennedy, J., Dayem, L., Carney, J. and Kealey, J. (2003)
Metabolic engineering of Escherichia coli for improved 6-deoxyerythronolide B production
Journal of Industrial Microbiology and Biotechnology 30, 500-509

Naeimpoor, F. and Mavituna, F. (2000)
Metabolic flux analysis in *Streptomyces coelicolor* under various nutrient limiting conditions
Metabolic Engineering 2, 140-148

Naumova, I., Shashkov, A., Tul'skaya, E., Streshinskaya, G., Kozlova, Y., Potekhina, N., Evtushenko, L. and Stackebrandt, E. (2001)
Cell wall teichoic acids: structural diversity, species specificity in the genus Nocardiopsis, and chemotaxonomic perspective
FEMS Microbiology Reviews 25, 269-283

Neidhardt, F., Ingraham, J., Low, K., Magasanik, B., Schaechter, M. and Umbarger, H. (1987)
Escherichia coli and *Salmonella typhimurium*
In Cellular and Molecular Biology
Published by The American Society for Microbiology.

Neumann, T., Piepersberg, W. and Distler, J. (1996)

Decision phase regulation of streptomycin production in *Streptomyces griseus*
Microbiology 142, 1953-1963

Nielsen, J. (1997)
Physiological engineering aspects of *Penicillium chrysogenum*
Singapore, World Scientific Publishing Co.

Olukoshi, E. and Packter, N. (1994)
Importance of stored triacylglycerols in *Streptomyces*: possible carbon source for antibiotics
Microbiology 140, 931-943

Oku, H., Fujita, K., Nomoto, T., Suzuki, K., Iwasaki, H. and Chinen, I. (1998)
NADH-dependent inhibition of branched-chain fatty acid synthesis in *Bacillus subtilis*
Bioscience, Biotechnology and Biochemistry 62, 622-627

Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M. (2001)
Genome sequence of an industrial microorganism *Streptomyces avermitilis*: Deducing the ability of producing secondary metabolites.
Proceedings of the National Academy of Science 98, 12215-12220

Omura, S., Tsuzuki, K., Tanaka, Y., Sakakibara, H., Aizawa, M. and Lukacs, G. (1983)
Valine as the precursor of *n*-butyrate unit in the biosynthesis of macrolide aglycone
The Journal of Antibiotics 36, 614-616

Omura, S., Taki, A., Kazuko, M. and Tanaka, Y. (1984)
Ammonium ions suppress the amino acid metabolism involved in the biosynthesis of protylonolide in a mutant of *Streptomyces fradiae*
The Journal of Antibiotics 37, 1362-1369

Overbeek, R., Larsen, N., Pusch, G., D'Souza, M., Selkov, E. Jr., Kyrpides, N., Fonstein, M., Maltsev, N. and Selkov, E. (2000)
WIT: integrated system for high-throughput genome sequence analysis and metabolic reconstruction
Nucleic Acids Research 28, 123-125

Overkamp, K., Bakker, B., Kötter, P., Luttik, M., van Dijken, J. and Pronk, J. (2002)
Metabolic engineering of glycerol production in *Saccharomyces cerevisiae*
Applied and Environmental Microbiology 68, 2814-2921

Pacey, M., Dirlam, J., Geldart, R., Leadlay, P., McArthur, H., McCormick, E., Monday R., O'Connell, T., Staunton, J. and Winchester, T. (1998)
Novel erythromycins from a recombinant *Saccharopolyspora erythraea* strain NRRL 2338 pIG1 - I. Fermentation, isolation and biological activity
Journal of Antibiotics 51, 1029-1034

Packter, N. and Olukoshi, E. (1995)
Ultrastructural studies of neutral lipid localisation in *Streptomyces*

Archives of Microbiology 164, 420-427

Padilla, G., Hindle, Z., Callis, R., Corner, A., Ludovice, M., Liras, P. and Baumberg, S. (1991)
The relationship between primary and secondary metabolism in *streptomyces*
In Genetics and Product Formation in *Streptomyces*, edited by Baumberg *et al.*, Plenum Press New York, 1991

Pai, S. and Kubitschek, H. (1992)
Catabolic pools in *Escherichia coli*
Research in Microbiology 143, 173-181

Palsson, B. (2000)
The challenges of in silico biology.
Nature Biotechnology 18, 1147-1150

Parsons, T. and Preiss, J. (1978)
Biosynthesis of bacterial glycogen
Journal of Biological Chemistry 253, 7638-7645

Petersen, S., de Graaf A., Eggeling, L., Möllney, M., Wiechert, W. and Sham, H. (2000)
In vivo quantification of parallel and bidirectional fluxes in the anaplerosis of *Corynebacterium glutamicum*
Journal of Biological Chemistry 275, 35932-35941

Pfeifer, B., Admiraal, S., Gramajo, H., Cane, D. and Khosla, C. (2001)
Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*
Science 291, 1790

Pfeifer, B., Hu, Z., Licari, P. and Khosla, C. (2002)
Process and metabolic strategies for improved production of *Escherichia coli*-derived 6-deoxyerythronolide B
Applied and Environmental Microbiology 68, 3287-3292

Phol, N., Gokhale, R., Cane, D. and Khosla, C. (1998)
Synthesis and incorporation of an N-acetylcysteamine analogue of methylmalonyl-CoA by a modular polyketide synthase
Journal of the American Chemical Society 120, 11206-11207

Pirt, S. (1965)
The maintenance energy of bacteria in growing cultures
Proceedings of the Royal Society of London. Series B, Biological Sciences 163, 224-231

Ranade, N. and Vining, L. (1993)
Accumulation of intracellular carbon reserves in relation to chloramphenicol biosynthesis by *Streptomyces venezuelae*
Canadian Journal of Microbiology 39, 377-383

Ranganathan, A., Timoney, M., Bycroft, M., Cortés, J., Thomas, I., Wilkinson, B., Kellenberger, L., Hanefeld, U., Galloway, I., Staunton, J. and Leadlay, P. (1999)
Knowledge-based design of bimodular and trimodular polyketide synthases based on domain and module swaps: a route to simple statin analogues
Chemistry & Biology 6, 731-741

Ratledge, C. and Wilkinson, S. (1988)
Microbial Lipids 1
Academic Press

Reeve, L. and Baumberg, S. (1998)
Physiological controls of erythromycin production by *Saccharopolyspora erythraea* are exerted at least in part at the level of transcription
Biotechnology Letters 20, 585-589

Reeves, A., Post, D. and Vanden Boom, T. (1998)
Physical-genetic map of the erythromycin-producing organism *Saccharopolyspora erythraea*
Microbiology 144, 2151-2159

Rodriguez, E., Hu, Z., Ou, S., Volchegursky, Y., Hutchinson, R. and McDaniel, R. (2003)
Rapid engineering of polyketide overproduction by gene transfer to industrially optimized strains
Journal of Industrial Microbiology and Biotechnology 30, 480-488

Rohr, J. (2000)
A new role for Polyketides
Angewandte Chemie International Edition 39, 2847-2849

Rose, I., Warms, J. and Kuo, D. (1992)
Proton transfer in catalysis by Fumarase
Biochemistry 31, 9993-9999

Roszkowski, J., Ruczaj, Z., Sawnor-Korszynska, D., Kotiuszko, D., Morawska, H., Siejko, D. and Raczynska-Bojanowska, K. (1971)
NADPH-regenerating systems in microorganisms producing macrolide antibiotics
Acta Microbiologica Polonica Series B 3, 97-106

Roubos, J., Krabben, P., Luiten, R., Babuska, R. and Heijnen, J. (2001)
A semi-stoichiometric model for a *Streptomyces* fed batch cultivation with multiple feeds. (D. Dochain and M. Perrier (Eds.)
Proceedings of the Compendium for Applied Biotechnology, 8. 299-304

Roubos, J. (2002)
Bioprocess modelling and optimization: Fed batch clavulanic acid production by *Streptomyces clavuligerus*
PhD thesis, Delft University of Technology

Rowe, C., Cortés, J., Gaisser, S., Staunton, J. and Leadlay, P. (1998)

Construction of new vectors for high-level expression in actinomycetes
Gene 216, 215-223

Savinell, J. and Palsson, B. (1992)
Optimal selection of metabolic fluxes for *in vivo* measurement. II. Application to *Escherichia coli* and hybridoma cell metabolism
Journal of Theoretical Biology 155, 215-242

Salah-Bey, K., Doumith, M., Michel, J., Haydock, S., Cortés, J., Leadlay, P. and Raynal, M. (1998)
Targeted gene inactivation for the elucidation of deoxysugar biosynthesis in the erythromycin producer *Saccharopolyspora erythraea*
Molecular and General Genetics 257, 542-553

Sapan, C., Lundblad, R. and Price, N. (1999)
Colorimetric protein assay techniques
Biotechnology and Applied Biochemistry 29, 99-108

Sauer, U., Hatzimanikatis, V., Bailey, J., Hochuli, M., Szyperski, T. and Wüthrich, K. (1997)
Metabolic fluxes in riboflavin-producing *Bacillus subtilis*
Nature Biotechnology 15, 448-449

Sauer, U. (2001)
Evolutionary engineering of industrially important microbial phenotypes
Advances in Biochemical Engineering and Biotechnology 73, Pages129-169

Scherman, M., Weston, A., Duncan, K., Whittington, A., Upton, R., Deng, L., Comber, R., Friedrich, J. and McNeil, M. (1995)
Biosynthetic origins of mycobacterial cell wall arabinosyl residues
Journal of Bacteriology 177, 7125-7130

Schneider, D., Bruton, C. and Charter, K. (2000)
Duplicated gene clusters suggest an interplay of glycogen and trehalose metabolism during sequential stages of aerial mycelium development in *Streptomyces coelicolor* A3(2)
Molecular and General Genetics 263, 543-553

Schuster, S., Fell, D. and Dandekar, T. (2000)
A general definition of metabolic pathways useful for systematic organisation and analysis of complex metabolic networks
Nature Biotechnology 18, 326-332

Seno, E. and Hutchinson, R. (1986)
The biosynthesis of tylosin and erythromycin: model systems for studies of the genetics and biochemistry of antibiotic formation
In "The Bacteria" 9, Published by Academic Press.

Seow, K., Meurer, G., Gerlitz, M., Wendt-pienkowski, E., Hutchinson, R. and Davies, J. (1997)

A study of iterative type II polyketide synthases using bacterial genes cloned from soil DNA: a means to access and use genes from uncultured microorganisms
Journal of Bacteriology 179, 7360-7368

Sequeira, S., Avignone-Rossa, C., Bushell, M., Butler, M. and Chater, K. (2003)
Distribution of metabolic fluxes in *Streptomyces coelicolor* strains with reduced pentose phosphate pathway activity leading to enhanced antibiotic production
Poster presentation at the *Streptomyces* Dissemination Meeting, University of Surrey, 22-23 July 2003

Shahab, N., Flett, F., Oliver, S. and Butler, P. (1996)
Growth rate control of protein and nucleic acid content in *Streptomyces coelicolor* A3(2) and *Escherichia coli* B/r
Microbiology 142, 1927-1935

Shen, B. (2003)
Polyketide biosynthesis beyond the type I, II, III, polyketide synthase paradigms
Current Opinion in Chemical Biology 7, 285-295

Shen, L. and Atkinson, D. (1970)
Regulation of pyruvate dehydrogenase from *Escherichia coli*: interaction of adenylate energy charge and other regulatory mechanisms
The Journal of Biological Chemistry 245, 5974-5978

Shim, M., Kim, W. and Kim, J. (1997)
Neutral lipids and lipase activity for actinorhodin biosynthesis of *Streptomyces coelicolor* A3(2)
Biotechnology Letters 19, 221-223

Silva, P., Bibi, F., De Le Paz Santangelo, M., Romano, M., Martín, C., Cataldi, A. and Aínsa, J. (2001)
Characterization of P55 a multidrug efflux pump in *Mycobacterium bovis* and *Mycobacterium tuberculosis*
Antimicrobial Agents and Chemotherapy 45, 800-804

Smith, D., Wood, N. and Hodgson, D. (1995)
Interaction between primary and secondary metabolism in *Streptomyces coelicolor* A3(2): role of pyrroline-5-carboxylate dehydrogenase
Microbiology 145, 1739-1744

Soga, T., Ueno, Y., Naraoka, H., Ohashi, Y., Tomita, M. and Nishioka, T. (2002)
Simultaneous determination of anionic intermediates for *Bacillus subtilis* metabolic pathways by capillary electrophoresis electrospray ionization mass spectrometry.
Analytical Chemistry 74, 2233-2239

Stafford, D. and Stephanopoulos, G. (2001)
Metabolic engineering as an integrating platform for strain development
Current Opinion in Microbiology 4, 336-340

Stanbury, P., Whitaker, A. and Hall, S. (1995)

Principles of fermentation technology, second edition
Pergamon Press

Stassi, D., Post, D., Satter, M., Jackson, M. and Maine, G. (1998)
A genetically engineered strain of *Saccharopolyspora erythraea* that produces 6,12-dideoxyerythromycin A as the major fermentation product
Applied Microbial Biotechnology 49, 725-731

Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S. and Gilles, E. (2002)
Metabolic network structure determines key aspects of functionality and regulation
Nature 420, 190-193

Stemmer, W. (1994)
Rapid evolution of a protein *in vitro* by DNA shuffling
Nature 370, 324-325

Stephanopoulos, G. and Valline, J. (1991)
Network rigidity and metabolic engineering in metabolite overproduction
Science 262, 1675-1681

Stephanopoulos, G., Aristidou, A. and Nielsen, J. (1998)
Metabolic engineering principles and methodologies
Academic Press

Stouthamer, A. (1979)
The search for correlation between theoretical and experimental growth yields.
In International Review of Biochemistry: Microbial Biochemistry 21, 1-47
Edited by J. Quayle. Baltimore: University Park Press

Stryer, L. (1995)
Biochemistry (fourth edition)
W. H. Freeman and Company

Summers, R., Donadio, S., Staver, M., Wendt-Pienkowski, E., Hutchinson, R. and Katz, L. (1997)
Sequencing and mutagenesis of genes from the erythromycin biosynthetic gene cluster of *Saccharopolyspora erythraea* that are involved in L-mycarose and D-desosamine production
Microbiology 143, 3251-3262

Surowitz, K. and Pfister, R. (1985)
Glucose metabolism and pyruvate excretion by *Streptomyces alboniger*
Canadian Journal of Microbiology 31, 702-706

Tang, L., Zhang, Y. and Hutchinson, R. (1994)
Amino acid catabolism and antibiotic synthesis: Valine is a source of precursors for macrolide biosynthesis in *Streptomyces ambofaciens* and *Streptomyces fradiae*
Journal of Bacteriology 176, 6107-6119

Tang, L., Fu, H. and McDaniel, R. (2000)

Formation of functional heterologous complexes using subunits from the picromycin, erythromycin and oleandomycin polyketide synthases
Chemistry & Biology 7, 77-84

Trefzer, A., Salas, J. and Bechthold, A. (1999)
Genes and enzymes involved in deoxysugar biosynthesis in bacteria
Natural Products Report 16, 283-299

Tsai, S. and Lee, Y. (1988)
Application of the metabolic pathway stoichiometry to statistical analysis of bioreactor measurement data
Biotechnology and Bioengineering 32, 713-715

Ushio, M. (2003)
A Metabolic Engineering Approach to Examine Polyketide Production by *Saccharopolyspora erythraea*
PhD Thesis, University of London

van Gulik, W., de Laat, W., Vinke, J. and Heijnen, J. (2000)
Application of Metabolic flux analysis for the identification of metabolic bottlenecks in the biosynthesis of penicillin-G
Biotechnology and Bioengineering 68, 602-618

Verma, S., Bhatia, Y., Padinhara, S. and Roy, I. (2002)
A possible role of poly-3-hydroxybutyric acid in antibiotic production in *Streptomyces*
Archives of Microbiology 179, 66-69

Voet, D. and Voet, J (1990)
Biochemistry
Publisher John Wiley and Sons

Waksman, S. (1919)
Cultural studies of species of *Actinomyces*
Soil Science 8, 71-215

Walker-Simmons, M. and Atkinson, D
Functional capacities and the adenylate energy charge in *Escherichia coli* under conditions of nutritional stress
Journal of Bacteriology 130, 676-683

Wallace, K., Zhao, B., McArthur, H. and Reynolds, K. (1995)
In vivo analysis of straight-chain and branched-chain fatty acid biosynthesis in three actinomycetes
FEMS Microbiology Letters 131, 227-234

Weissman, K., Timoney, M., Bycroft, M., Grice, P., Hanefeld, U., Staunton, J. and Leadlay, P. (1997)
The molecular basis of Celmer's rules: The stereochemistry of the condensation step in chain extension on the erythromycin polyketide synthase
Biochemistry 36, 13849-13855

Weitzman, P. (1972)
Regulation of α-ketoglutarate dehydrogenase activity in *Acinetobacter*
FEBS Letters 22, 323-326

Weymouth-Wilson, A. (1997)
The role of carbohydrates in biologically active natural products
Natural Product Reports 14, 99-110

Woodrow, C., Burchmore, R. and Krishna, S. (2000)
Hexose permeation pathways in Plasmodium falciparum-infected erythrocytes
Proceedings of the National Academy of Sciences 97, 9931-9936

Zhao, L., Sherman, D. and Liu, H. (1998a)
Biosynthesis of Desosamine: Molecular evidence suggesting β-glucosylation as a self-resistance mechanism in methymycin/neomthymycin producing strain, *Streptomyces venszuelae*
Journal of the American Chemical Society 120, 9374-9375

Zhao, L., Sherman, D. and Liu, H. (1998b)
Biosynthesis of Desosamine: Construction of a new methymycin/neomethymycin analogue by deletion of a desosamine biosynthetic gene
Journal of the American Chemical Society 120, 10256-10257

Zhou, Z., Liu, Z., Qian, Y., Kin, S. and Goodfellow, M. (1998)
Saccharopolyspora spinosporotrichia sp. Nov., a novel actinomycete from soil
International Journal of Systematic Bacteriology 48, 53-58

Zöllner, N. and Kirsch, K. (1962)
Über die quantitative bestimmung von lipoiden (mikromethode) mittels der vielen nat ürlichen lipoiden (allen bekannten plasmalipoiden) gemeinsamen sulfophosphovanillin-reaktion
Zeitschrift für die Gesamte Experimentelle Medizin 135, 545-561

Zor, T. and Selinger, Z. (1996)
Linearization of the Bradford protein assay increases its sensitivity: Theoretical and experimental studies
Analytical Biochemistry 236, 302-308

9 APPENDICES

9.1 Appendix 1 Metabolic Network

Table 16 contains all the reactions in the metabolic network. The reaction name as used in the model and given in Figure 7 is given in the first column. Abbreviations used in the model are listed in Table 17. Some of the reactions are denoted by numbers and some by names relevant to the reactions. The numbered reactions are reactions inherited from the models of Daae (1999) and Ushio (2003). This nomenclature was retained for these reactions to facilitate comparisons. Reactions with descriptive names have been added during the course of this work. Descriptive names are easier to use in Flux Analyser as it refers to reactions in the command window by name. With numbered reactions this means that a way of translating these into reactions is needed before the results can be fully comprehended. The problem does not occur with the main user interface which has a visual representation of the network.

The way of writing biochemical reactions used in flux balance analysis is a little unusual. Many of the reactions appear unbalanced with respect to cofactors eg reaction 1 where ATP is hydrolysed but no ADP produced. This is because recycling of cofactors causes linear dependencies making modelling impossible. Cofactors are instead considered to be created and destroyed. So in reaction 1 in the forward direction 1 ATP is destroyed. In reaction 13 in the forward direction 1 ATP is created. This allows production and consumption of these cofactors to be balanced which if the steady state assumption holds true is a valid way of dealing with the problem.

Reactions which involve taking up substrates or secreting products are also a little unusual. Only metabolites within the cell are considered in the model. Molecules which cross the cell boundary are only considered on the inside. Eg in reaction 1 glucose is taken up and phosphorylated. The glucose on the outside of the cell is not considered in the model so the G6P seems to have come from nowhere. Likewise with reaction 3 once the α-ketoglutarate leaves the cell it is no longer considered by the model and so is not included on the right hand side of the equation.

Table 16: List of reactions contained in the stoichiometric matrix

Reaction Name	Reaction
1	1 ATP = 1 G6P
2	4 NADPH = 1 NH4
3	1 AKG =
4	1 PYR =
5	1 FUM =
6	1 CO2 =
7	2 G1P + 1 GLU + 9 NADPH + 2 O2 + 1 PCoA + 6 SMMCoA + 3 SAM + 2 dTTP = 1 AKG + 3 SAH + 2 dTDP
8	= 1 O2
9	1 G6P = 1 F6P
10	1 ATP + 1 F6P = 2 GAP
10b	2 GAP = 1 F6P
11	1 GAP = 1 ATP + 1 NADH + 1 PG
12	1 PG = 1 PEP
13	1 PEP = 1 ATP + 1 PYR
14	1 CO2 + 1 PEP = 1 OAA
15	1 ICIT = 1 SUC + 1 Glyx
16	1 AcCoA + 1 Glyx = 1 MAL
17	1 PYR = 1 AcCoA + 1 CO2 + 1 NADH
19	1 AcCoA + 1 OAA = 1 ICIT
20	1 ICIT = 1 AKG + 1 CO2 + 1 NADH
21	1 AKG = 1 CO2 + 1 NADH + 1 SucCoA
22	1 SucCoA = 1 ATP + 1 SUC
23	1 SUC = 1 FADH2 + 1 FUM
24	1 FUM = 1 MAL
25	1 MAL = 1 NADH + 1 OAA
26	1 G6P = 1 CO2 + 1 NADH + 1 NADPH + 1 Ru5P
27	1 Ru5P = 1 X5P
28	1 Ru5P = 1 R5P
30	1 R5P + 1 X5P = 1 E4P + 1 F6P
31	1 E4P + 1 X5P = 1 F6P + 1 GAP
32	1 GLU + 1 PYR = 1 AKG + 1 ALA
33	1 GLU + 1 NADPH + 2 PYR = 1 AKG + 1 CO2 + 1 VAL
34	1 AcCoA + 1 GLU + 1 NADPH + 2 PYR = 1 AKG + 2 CO2 + 1 LEU + 1 NADH
35	1 GLU + 1 OAA = 1 AKG + 1 ASP
36	1 ASP + 2 ATP + 1 NH4 = 1 ASN
39	1 ASP + 2 ATP + 2 NADPH = 1 THR
40	1 GLU + 1 NADPH + 1 PYR + 1 THR = 1 AKG + 1 CO2 + 1 ILE + 1 NH4
41	1 AKG + 1 NADPH + 1 NH4 = 1 GLU
42	1 AKG + 1 GLN + 1 NADPH = 2 GLU
43	1 ATP + 1 GLU + 1 NH4 = 1 GLN
44	1 AcCoA + 1 ATP + 2 GLU + 1 NADPH = 1 AKG + 1 ORN + 1 ACT
45	1 ATP + 1 GLU + 1 NADH + 1 NADPH = 1 PRO
46	1 GLU + 1 PG = 1 AKG + 1 NADH + 1 SER
47	1 SER = 1 GLY + 1 MTHF
48	1 AcCoA + 3 ATP + 4 NADPH + 1 SER = 1 CYS + 1 ACT

49	1 ATP + 1 E4P + 2 PEP = 1 CHOR + 1 NADH
50	1 CHOR + 1 GLU = 1 AKG + 1 CO2 + 1 PHE
51	1 CHOR + 1 GLU = 1 AKG + 1 CO2 + 1 NADH + 1 TYR
52	1 CHOR + 1 GLN + 1 PRPP + 1 SER = 1 CO2 + 1 GAP + 1 GLU + 1 PYR + 1 TRP
53	1 ATP + 1 GLN + 1 PRPP = 1 AKG + 1 HIS + 2 NADH + 1 AICAR
54	2 ATP + 1 R5P = 1 PRPP
55	1 ASP + 4 ATP + 1 CO2 + 1 FTHF + 2 GLN + 1 GLY + 1 PRPP = 1 FUM + 2 GLU + 1 AICAR
56	1 FTHF + 1 AICAR = 1 IMP
57	1 ASP + 1 ATP + 1 IMP = 1 FUM + 1 AMP
58	1 ATP + 1 GLN + 1 IMP = 1 GLU + 1 NADH + 1 GMP
59	1 ASP + 1 PRPP + 1 CBP = 1 CO2 + 1 NADH + 1 UMP
60	1 ATP + 1 GLN + 1 UMP = 1 CMP + 1 GLU
61	1 ATP + 1 MTHF + 2 NADPH + 1 UMP = 1 dTDP
62 (Used in flux balance analysis	1 AcCoA + 1 ATP + 1 NADPH = 1 CO2 + 1 FADH2 + 1 FTHF + 1 NADH
62 (Used in elementary mode analysis	1 CO2 + 1ATP + 1 NADPH = 1 FTHF
63	1 FTHF + 1 NADPH = 1 MTHF
64	2 NADH + 1 O2 = 4 ATP
65	2 FADH2 + 1 O2 = 2 ATP
67	1 ATP =
68	1 SucCoA = 1 RMMCoA
69	1 RMMCoA = 1 SMMCoA
70	1 RMMCoA = 1 CO2 + 1 PCoA
72	1 G6P = 1 G1P
Acetate	2 ATP + 1 ACT = 1 AcCoA
Acetoacetate_metabolism	2 AcCoA = 1 AcAcCoA
ADPglc_synth	2 ATP + 1 G1P = 1 ADPglc
AMP_deg	1 AMP = 1 NADH + 1 NH4 + 1 R5P + 1 XAN
ARG_deg	1 AKG + 1 ARG + 1 O2 = 1 CO2 + 1 GLU + 1 NADH + 1 NH4 + 1 SUC + 1 UREA
Asparagine_catabolism	1 ASN = 1 ASP + 1 NH4
Carbamoyl_phosphate_synth	2 ATP + 1 CO2 + 1 GLN = 1 GLU + 1 CBP
CMP_deg	1 AKG + 1 CMP + 1 NADPH = 1 CO2 + 1 GLU + 1 NADH + 2 NH4 + 1 R5P + 1 RMMCoA
CTP_synth	2 ATP + 1 CMP = 1 CTP
Cysteine_Catabolism	1 CYS = 1 NH4 + 1 PYR
dATP_synth	1 NADPH + 1 rATP = 1 dATP
dCTP_synth	1 NADPH + 1 CTP = 1 dCTP
dGTP_synth	1 NADPH + 1 GTP = 1 dGTP
DPPG_synth	2 CTP + 3 G3P + 4 FA = 2 CMP + 1 DPPG
dTTP_synth	1 ATP + 1 dTDP = 1 dTTP
ECP	0.137 ALA + 0.0837 ARG + 0.017 ASN + 0.0613 ASP + 4.27 ATP + 0.00775 CYS + 0.0263 GLN + 0.0569 GLU + 0.0966 GLY + 0.0234 HIS + 0.0287 ILE + 0.102 LEU + 0.0205 LYS + 0.0158 MET + 0.0265 PHE + 0.062 PRO + 0.0497 SER + 0.0616 THR + 0.0152 TRP + 0.0205 TYR + 0.0868 VAL =
Fatty_acid_synth	0.048 AcCoA + 0.669 ATP + 12.728 NADPH + 0.71 PCoA + 0.49 AKMV + 0.421 AKV + 5.695 MALCoA = 6.854 CO2 + 1 FA
GL_synth	1.5 G6P + 1.5 CTP + 1 G3P + 2 FA = 1.5 CMP + 1 GL

Glycerol1	1 ATP = 1 G3P
Glycerol2	1 GAP + 1 NADPH = 1 G3P
Glycine_catabolism	1 GLY = 1 CO2 + 1 MTHF + 1 NADH + 1 NH4
GMP_deg	1 GMP = 1 NH4 + 1 R5P + 1 XAN
GTP_synth	2 ATP + 1 GMP = 1 GTP
Histidine_catabolism	1 HIS = 1 GLU + 1 MTHF + 2 NH4
Isoleucine_catabolism_1	1 AKG + 1 ILE = 1 GLU + 1 AKMV
Isoleucine_catabolism_2	1 AKMV = 1 AcCoA + 1 CO2 + 1 FADH2 + 1 NADH + 1 PCoA
Leucine_deg	1 AKG + 1 ATP + 1 LEU + 1 SucCoA = 2 CO2 + 1 FADH2 + 1 GLU + 1 NADH + 1 SUC + 2 AcAcCoA
Lysine_biosyntheis1	1 ASP + 1 ATP + 1 GLU + 2 NADPH + 1 PYR + 1 SucCoA = 1 AKG + 1 SUC + 1 mDAP
Lysine_biosynthesis_2	1 mDAP = 1 CO2 + 1 LYS
Lysine_degredation	2 AKG + 1 LYS = 1 CO2 + 2 GLU + 1 GlutCoA
Malate_shunt	1 MAL = 1 CO2 + 1 NADPH + 1 PYR
MALCoA_synth	1 AcCoA + 1 ATP + 1 CO2 = 1 MALCoA
Methionine_biosynthesis1	1 ASP + 1 ATP + 1 CYS + 2 NADPH + 1 SucCoA = 1 SUC + 1 CTT
Methionine_biosynthesis2	1 CTT = 1 NH4 + 1 PYR + 1 HCYS
Methionine_biosynthesis3	1 MTHF + 1 NADH + 1 HCYS = 1 MET
NH4_excretion	1 NH4 =
PEP_C_K	1 ATP + 1 OAA = 1 CO2 + 1 PEP
PHE_deg	1 O2 + 1 PHE = 1 TYR
PPC_synth	1 SER + 3 SAM + 1 CTP + 1 G3P + 2 FA = 1 CMP + 1 CO2 + 3 SAH + 1 PPC
PPE_synth	1 SER + 1 CTP + 1 G3P + 2 FA = 1 CMP + 1 CO2 + 1 PPE
PPI_synth	1 G6P + 1 CTP + 1 G3P + 2 FA = 1 CMP + 1 PPI
PPM_synth	1 SER + 1 SAM + 1 CTP + 1 G3P + 2 FA = 1 CMP + 1 CO2 + 1 SAH + 1 PPM
Proline_catabolism	0.5 O2 + 1 PRO = 1 GLU + 1 NADH
Propionyl_MMCoA_swapping	1 CO2 + 1 PCoA = 1 SMMCoA
rATP_synth	2 ATP + 1 AMP = 1 rATP
RED	1 F6P + 3 NADPH + 2 O2 + 1 dTTP + 5 MALCoA = 5 CO2 + 1 dTDP
Sam_meatbolism1	1 ATP + 1 MET = 1 SAM
SAM_metabolism2	2 ATP + 1 SAH = 1 HCYS
Serine_catabolism	1 SER = 1 NH4 + 1 PYR
Stat_Ara+	1 UDPara = 1 UMP
Stat_DNA	1.515 ATP + 0.115 dTTP + 0.115 dATP + 0.385 dGTP + 0.385 dCTP =
stat_Gal+	1 UDPgal = 1 UMP
Stat_Glyc	1 ADPglc =
Stat_Lip-	1 ATP = 17.229 AcCoA + 19.092 FADH2 + 19.092 NADH + 2.13 PCoA + 1.47 AKMV + 1.263 AKV + 1 G3P
Stat_Lip+	0.656 TAG + 0.027 DPPG + 0.053 PPE + 0.052 PPM + 0.051 PPC + 0.046 PPI + 0.13 GL =
Stat_Pptg	2 ALA + 5 ATP + 1 GLU + 1 GLY + 1 mDAP + 1 UDPNAG + 1 UDPNAM =
Stat_Prot_-	= 0.137 ALA + 0.0833 ARG + 0.017 ASN + 0.0613 ASP + 0.00775 CYS + 0.0263 GLN + 0.0569 GLU + 0.0967 GLY + 0.0234 HIS + 0.0286 ILE + 0.103 LEU + 0.0205 LYS + 0.0158 MET + 0.0265 PHE + 0.0621 PRO + 0.0497 SER + 0.0616 THR + 0.0152 TRP + 0.0205 TYR + 0.0868 VAL
Stat_Prot_+	0.137 ALA + 0.0833 ARG + 0.017 ASN + 0.0613 ASP + 4.289 ATP + 0.00775 CYS + 0.0263 GLN + 0.0569 GLU + 0.0967

	GLY + 0.0234 HIS + 0.0286 ILE + 0.103 LEU + 0.0205 LYS + 0.0158 MET + 0.0265 PHE + 0.0621 PRO + 0.0497 SER + 0.0616 THR + 0.0152 TRP + 0.0205 TYR + 0.0868 VAL =
Stat_RNA	0.4 ATP + 0.189 UTP + 0.252 CTP + 0.341 GTP + 0.22 rATP =
Stat_RNA-	= 0.252 CMP + 0.189 UMP + 0.22 AMP + 0.341 GMP
Stat_tre-	2 ATP = 2 G6P
Stat_tre+	1 G6P + 1 UDPglc = 1 UMP
TAG_synth	1 G3P + 3 FA = 1 TAG
Threonine_catabolism	1 THR = 1 CO2 + 1 NADH + 1 NH4 + 1 PCoA
Trans-sulphuration1	1 SER + 1 HCYS = 1 CTT
Trans-sulphuration2	1 CTT = 1 CO2 + 1 CYS + 1 NADH + 1 NH4 + 1 PCoA
Tryptophan_catabolism_1	1 NADPH + 3 O2 + 1 TRP = 1 ALA + 1 CO2 + 1 ACMS
Tryptophan_catabolism2.1	1 ACMS = 1 CO2 + 1 NADH + 1 GlutCoA
Tryptophan_catabolism2.2	1 NADPH + 1 ACMS = 1 CO2 + 2 NADH + 1 GlutCoA
Tryptophan_catabolism3	1 GlutCoA = 1 CO2 + 1 FADH2 + 1 NADH + 1 AcAcCoA
TYR_deg	1 AKG + 2 O2 + 1 SucCoA + 1 TYR = 1 CO2 + 1 FUM + 1 GLU + 1 SUC + 1 AcAcCoA
UDPara_synth	1 ATP + 1 Ru5P + 1 UTP = 1 UDPara
UDPgal_synth	1 G1P + 1 UTP = 1 UDPgal
UDPglc_synth	1 G1P + 1 UTP = 1 UDPglc
UDPNAG_formation	1 AcCoA + 1 F6P + 1 GLN + 1 UTP = 1 GLU + 1 UMP + 1 UDPNAG
UDPNAM_synth	1 NADPH + 1 PEP + 1 UDPNAG = 1 UDPNAM
UMP_deg	1 AKG + 1 NADPH + 1 UMP = 1 CO2 + 1 GLU + 1 NADH + 1 NH4 + 1 R5P + 1 MALCoA
Urea_cycle_1	1 ASP + 2 ATP + 1 CBP + 1 ORN = 1 ARG + 1 FUM
Urea_cycle_2	1 ARG = 1 ORN + 1 UREA
UREA_deg	1 UREA = 2 CO2 + 4 NH4
UTP_synth	2 ATP + 1 UMP = 1 UTP
Valine_catabolism_1	1 AKG + 1 VAL = 1 GLU + 1 AKV
Valine_catabolism_2	1 AKV = 2 CO2 + 1 FADH2 + 3 NADH + 1 PCoA
XAN_deg	0.5 O2 + 1 XAN = 1 CO2 + 1 NADH + 1 UREA + 1 Glyx

Note two versions of reaction 62 are given, one used for flux balance analysis and one used in elementatry modes analysis (see section 4.2.1.2).

Table 17: List of abbreviations of compounds in the metabolic network

Abbreviation	Name of Metabolite
AcAcCoA	Acetoacetyl-CoA
AcCoA	Acetyl-CoA
ACMS	2-Amino-3-carboxymuconate-6-semialdehyde
ACT	Acetic acid
ADPglc	ADP-glucose
AICAR	5-Aminoimidazole-4-carboxamide ribotide
AKG	α-Ketoglutarate
AKMV	α-Keto-β-methylvalerate
AKV	α-Ketoisovalerate
ALA	Alanine
AMP	Adenosine monophosphate
ARG	Arginine
ASN	Asparagine
ASP	Aspartic-acid

ATP	Adenosine triphosphate
CBP	Carbamoyl phosphate
CHOR	Chorismate
CMP	Cytosine monophosphate
CO2	Carbon dioxide
CTP	Cytidine diphosphate
CTT	Cystathionine
CYS	Cysteine
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytosine triphosphate
dGTP	Deoxyguanosine triphosphate
DPPG	Diphosphatidylglycerol
dTDP	Deoxythymidine diphosphate
dTTP	Deoxythymidine triphosphate
E4P	Erythrose-4-phosphate
F6P	Fructose-6-phosphate
FA	Fatty acid
FADH2	Flavin adenine dinucleotide (reduced form)
FTHF	Formyl-tetrahydrofolate
FUM	Fumarate
G1P	Glucose-1-phosphate
G3P	Glycerol-3-phosphate
G6P	Glucose-6-P
GAP	Glyceraldehyde-3-phosphate
GL	Glycolipid
GLN	Glutamine
GLU	Glutamate
GlutCoA	Glutaryl-CoA
GLY	Glycine
Glyx	Glyoxylate
GMP	Guanosine monophosphate
GTP	Guanosine triphosphate
HCYS	Homocysteine
HIS	Histidine
ICIT	Isocitrate
ILE	Isoleucine
IMP	Inosine-monophosphate
LEU	Leucine
LYS	Lysine
MAL	Malate
MALCoA	Malonyl-CoA
mDAP	*meso*-α,ε-Diaminopimelate
MET	Methionine
MTHF	Methyl-tetrahydrofolate
NADH	Nicotinamide adenine dinucleotide (reduced form)
NADPH	Nicotinamide adenine dinucleotide phosphate (reduced form)
NH4	Ammonium
O2	Oxygen
OAA	Oxaloacetate
ORN	Ornithine
PCoA	Propionyl-CoA
PEP	Phosphoenolpyruvate

PG	Phosphoglycerate
PHE	Phenylalanine
PPC	Phosphatidylcholine
PPE	Phosphatidylethanolamine
PPI	Phosphatidylinositol
PPM	Phosphatidylmonomethylethanolamine
PRO	Proline
PRPP	Phosphoribosyl-pyrophosphate
PYR	Pyruvate
R5P	Ribulose-5-phosphate
rATP	ATP
RMMCoA	R-Methylmalonyl-CoA
Ru5P	Ribulose-5-phosphate
SAH	*S*-Adenosylhomocysteine
SAM	*S*-Adenosylmethionine
SER	Serine
SMMCoA	S-Methylmalonyl-CoA
SUC	Succinate
SucCoA	Succinyl-CoA
TAG	Triacylglycerol
THR	Threonine
TRP	Tryptophan
TYR	Tyrosine
UDPara	UDP-arabinose
UDPgal	UDP-galactose
UDPglc	UDP-glucose
UDPNAG	UDP-*N*-acetylglucosamine
UDPNAM	UDP-*N*-Acetylmuramic Acid
UMP	Uridine monophosphate
UREA	Urea
UTP	Uridine triphosphate
VAL	Valine
X5P	Xyulose-5phosphate
XAN	Xanthine